WARSAW UNIVERSITY OF TECHNOLOGY

Ph.D. Thesis

Discipline of Science: Mathematics/
Field of Science: Natural Sciences

Wojciech Wójciak, MSc

Multi-Domain Optimum Sample Allocation with Controlled-Precision under Upper-Bound Constraints

Supervisor
Professor **Jacek Wesołowski,** PhD, DSc

Warsaw 2026

To my parents

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Professor Jacek Wesołowski, for inspiring me to undertake this work, for his valuable suggestions and insights throughout the course of this research, and for his thoughtful guidance at every stage of this dissertation.

Abstract

Optimum sample allocation in stratified sampling is one of the fundamental issues in survey methodology. In its classical formulation, it involves determining how a given total sample size should be allocated among strata so that certain criteria, typically related to minimizing the variances of estimators, are satisfied.

This thesis addresses the multi-domain optimum sample allocation problem in stratified sampling. The objective is the simultaneous minimization of the variance of the global total estimator and the variances of domain total estimators under a fixed total sample size, while ensuring that allocations do not exceed the population sizes within strata. The problem has been formulated in a way that allows controlling the relative precision of estimators across domains through pre-specified domain-wise priority weights. A review of the available literature indicates that no exact analytical algorithm has yet been developed to solve this problem.

The main contributions of this thesis are twofold. First, it formulates sufficient optimality conditions for the solution to the allocation problem under consideration. Second, it defines the RDCA algorithm, which solves this allocation problem based on the established conditions. The optimality conditions were derived using convex optimization techniques (e.g., Karush-Kuhn-Tucker conditions) and matrix algebra methods, with analyses involving the eigenvalue problem. A proof of correctness of the proposed algorithm constitutes an essential part of the thesis.

An integral part of this dissertation is the implementation of the RDCA algorithm as a function in the R programming language, provided in the **stratallo** package, which is publicly available in the Comprehensive R Archive Network.

Keywords: optimum sample allocation, multi-domain allocation, variance minimization, controlled-precision, stratified sampling, exact algorithm, RDCA algorithm, Karush-Kuhn-Tucker conditions, eigenvalues, eigenvectors.

Streszczenie

Alokacja optymalna próby w warstwowym schemacie próbkowania jest jednym z podstawowych zagadnień metodologii badań reprezentacyjnych. W klasycznej postaci zagadnienie to polega na wyznaczeniu liczebności prób w warstwach przy ustalonej liczebności całkowitej próby, z uwzględnieniem kryteriów optymalności związanych z minimalizacją wariancji estymatorów.

Niniejsza praca koncentruje się na problemie wielodomenowej alokacji optymalnej próby w schemacie warstwowym. Celem jest jednoczesna minimalizacja wariancji estymatora sumy globalnej oraz wariancji estymatorów sum w poszczególnych domenach, przy ustalonej całkowitej liczebności próby i zapewnieniu, że przydziały nie przekraczają liczebności populacji w warstwach. Problem ten sformułowano w sposób umożliwiający kontrolę względnej precyzji estymatorów w poszczególnych domenach za pomocą uprzednio określonych wag priorytetowych. Z dostępnej literatury wynika, że dotychczas nie opracowano dokładnego algorytmu analitycznego, który rozwiązywałby ten problem.

Główne rezultaty pracy obejmują sformułowanie warunków dostatecznych dla rozwiązania optymalnego rozważanego problemu alokacji oraz zdefiniowanie algorytmu RDCA, który rozwiązuje ten problem w oparciu o ustanowione warunki optymalności. Warunki optymalności wyprowadzono przy użyciu metod i narzędzi optymalizacji wypukłej (w tym warunków Karusha-Kuhna-Tuckera) oraz algebry macierzy w kontekście zagadnienia ich wartości własnych. Dowód poprawności zaproponowanego algorytmu stanowi istotny element pracy.

Integralną częścią dysertacji jest implementacja algorytmu RDCA jako funkcji w języku R, zawartej w pakiecie **stratallo**, dostępnym publicznie w repozytorium Comprehensive R Archive Network.

Słowa kluczowe: alokacja optymalna próby, alokacja wielodomenowa, minimalizacja wariancji, kontrolowana precyzja, próbkowanie warstwowe, dokładny algorytm, algorytm RDCA, warunki Karusha-Kuhna-Tuckera, wartości własne, wektory własne.

Contents

Chapter 1. Introduction 13

1.1. Preliminary Problem Formulation 14

1.2. Related Work 18

1.3. Motivation 20

1.3.1. Motivation for studying the problem 20

1.3.2. Motivation for an exact analytical algorithm – limitations of general-purpose NLP solvers 20

1.3.3. Remark on allocations that are not necessarily integer-valued 21

1.4. Main Results 22

1.5. Organization of the Thesis 23

Chapter 2. Population - Total Sample Size Model 25

2.1. Points Indexed by $\mathbb{N}^2$ 25

2.2. Model Definition 26

2.3. Auxiliary Functions 28

2.4. Sums and Cartesian Products over the Set of Domain-Stratum Indices $\mathcal{H}$ 31

Chapter 3. The Controlled-Precision Domain Allocation Problem 33

3.1. Problem Definition and Solvability 33

3.2. Optimality Conditions 40

3.2.1. The Eigen operator and the function s 40

3.2.2. A relaxed auxiliary problem and its equivalence 43

3.2.3. Optimality conditions 47

Chapter 4. The RDCA Algorithm 55

4.1. Algorithm Definition 56

4.2. Examples . . . 58
4.2.1. Two domains . . . 59
4.2.2. Three domains . . . 61
4.3. Implementation in R . . . 62

Chapter 5. Structural Properties of the RDCA Algorithm . . . 65

5.1. Notational Conventions . . . 66
5.2. Definiteness and Termination . . . 68
5.3. Basic Relations Between Program Variables . . . 74
5.4. The Last-Branch Recursion Path . . . 75
5.5. The take-max Strata Sets $\mathcal{V}_i$ . . . 82

Chapter 6. Correctness of the RDCA Algorithm . . . 87

6.1. The Form of the Variable $(T, \boldsymbol{x})$ . . . 89
6.2. Domain Blockage . . . 95
6.3. Monotonicity of the Function s . . . 101
6.4. Partial and Total Correctness . . . 115

Chapter 7. Discussion of the Results and Future Work . . . 117

7.1. Summary of Results . . . 117
7.2. Directions for Future Research . . . 118
7.2.1. New algorithms . . . 118
7.2.2. Generalizations of the CPDA problem . . . 118

Appendix A. List of Symbols . . . 121

Appendix B. List of Abbreviations . . . 123

Appendix C. Perron-Frobenius Theory . . . 125

Appendix D. Convex Sets and Convex Functions . . . 127

Appendix E. Selected Results in Topology . . . 129

Appendix F. Elements of Mathematical Optimization . . . 135

Appendix G. Multi-Domain Optimum Sample Allocation with Controlled-Precision without Upper-Bound Constraints . . . 139

Bibliography . . . 143

Chapter 1

Introduction

Survey sampling is a recognized area of statistics concerned with the design of sampling schemes and inference procedures for finite populations. It underpins a wide range of applications, including official statistics, the social sciences, and various applied disciplines. Foundational treatments of survey sampling theory and methods can be found in the classical books by Cochran [9], Lohr [26], and Särndal, Swensson and Wretman [42].

This thesis is concerned with an optimum sample allocation problem arising in a multi-domain survey setting. We briefly introduce two key concepts from survey sampling methodology that are essential to this problem: *multi-domain estimation* and *optimum sample allocation* in stratified sampling.

In many survey applications, estimates are required not only for the population as a whole but also for specific subpopulations. For example, in a national health survey, policymakers may require reliable estimates of disease prevalence not only at the national level but also for individual provinces or administrative regions. Such subpopulations, for which separate estimates are required, are referred to as *domains*. The concept of domain-level estimation is discussed in detail, for example, in Särndal et al. [42, Chapter 10, p. 386].

From the perspective of estimation and sampling design, one of the central topics in survey methodology is that of *optimum sample allocation*. This topic arises in the context of stratified sampling – a widely used sampling technique in which the population (or each domain) is divided into nonoverlapping subgroups called *strata*, from which probability samples are drawn. The problem of optimum sample allocation consists in determining the sample sizes for each stratum in such a way as to optimize a specified criterion. This criterion may involve, for example, minimizing the variance of an estimator or minimizing

the total cost of conducting a survey. The problem of optimum sample allocation in stratified sampling has a long tradition in the survey sampling literature; see, e.g., Cochran [9, Sec. 5.5, p. 96] and Särndal et al. [42, Sec. 3.7, p. 100]. It is worth noting that, in the optimum allocation problem, it is typically assumed that the strata have already been formed. This assumption is explicitly adopted in the allocation problem studied in this thesis. The formation of strata constitutes a distinct methodological problem and is discussed, for example, in Kish [23, Sec. 3.6, p. 98] and, more recently, in Lednicki and Wieczorkowski [25].

This dissertation is devoted to the problem of domain-wise optimum sample allocation. Specifically, we study a setting in which the population is divided into domains, each of which is further stratified. The objective is to determine the optimum allocation of a fixed total sample size across all strata within all domains so as to minimize the total variance of the resulting estimators. The optimization is carried out subject to constraints ensuring that the allocated sample size in each stratum does not exceed the stratum's population size and that certain domain-level estimation precision requirements are satisfied. We provide a precise formulation of this problem in the following Section 1.1.

1.1. Preliminary Problem Formulation

Consider a finite population consisting of N elements, denoted by $U = \{1, \ldots, N\}$. Let U be partitioned into D nonempty, disjoint subsets $U_1, \ldots, U_D$, called domains; that is,

$$U = \bigcup_{d=1}^{D} U_d. \tag{1.1}$$

Suppose that the parameter of interest is the total t_d of a univariate, positive-valued study variable y in domain U_d;

$$t_d = \sum_{k \in U_d} y_k, \qquad d \in \{1, \ldots, D\}, \tag{1.2}$$

where $y_k \in \mathbb{R}_+$ denotes the value of y for population element $k \in U$.

To estimate t_d for each $d \in \{1, \ldots, D\}$, we consider the π *estimator* (also known as the Horvitz-Thompson estimator) under a *stratified simple random sampling without replacement* design, hereafter abbreviated as the STSI design. This design combines *stratified sampling* (ST sampling) with *simple random sampling without replacement* (SI sampling), as described below.

Under ST sampling, each domain U_d, $d \in \{1, \ldots, D\}$, is partitioned into nonempty, disjoint strata,

$$U_d = \bigcup_{h=1}^{H_d} U_{d,h}, \tag{1.3}$$

where $U_{d,h}$ denotes the hth stratum within domain U_d, and H_d is the total number of strata in U_d. The size of stratum $U_{d,h}$ is denoted by $N_{d,h}$. For each $d \in \{1, \ldots, D\}$ and $h \in \{1, \ldots, H_d\}$, a probability sample $s_{d,h}$ of size $n_{d,h}$ is selected from stratum $U_{d,h}$ according to a specified sampling design. We assume that each stratum is sampled using SI sampling and that all samples $s_{d,h}$, $h \in \{1, \ldots, H_d\}$, $d \in \{1, \ldots, D\}$, are independent. Consequently, the overall sampling design is the STSI design.

Under the STSI design, the π estimator of t_d is given by

$$\hat{t}_d = \sum_{h=1}^{H_d} \frac{N_{d,h}}{n_{d,h}} \sum_{k \in s_{d,h}} y_k, \qquad d \in \{1, \ldots, D\}. \tag{1.4}$$

The variance of $\hat{t}_d$ is

$$\mathrm{Var}(\hat{t}_d) = \sum_{h=1}^{H_d} \Big(\frac{1}{n_{d,h}} - \frac{1}{N_{d,h}} \Big) N_{d,h}^2 S_{d,h}^2, \qquad d \in \{1, \ldots, D\}, \tag{1.5}$$

where

$$S_{d,h}^2 := \frac{1}{N_{d,h} - 1} \sum_{k \in U_{d,h}} \big(y_k - \bar{y}_{U_{d,h}} \big)^2, \tag{1.6}$$

is the population variance of the study variable y in stratum $U_{d,h}$, and $\bar{y}_{U_{d,h}} := \frac{1}{N_{d,h}} \sum_{k \in U_{d,h}} y_k$. It can also be shown that the overall variance of the π estimator

$$\hat{t} = \sum_{d=1}^{D} \sum_{h=1}^{H_d} \frac{N_{d,h}}{n_{d,h}} \sum_{k \in s_{d,h}} y_k, \tag{1.7}$$

of the total $t = \sum_{k \in U} y_k$ is

$$\mathrm{Var}(\hat{t}) = \sum_{d=1}^{D} \sum_{h=1}^{H_d} \Big(\frac{1}{n_{d,h}} - \frac{1}{N_{d,h}} \Big) N_{d,h}^2 S_{d,h}^2. \tag{1.8}$$

For more details on the STSI design and related estimators, see Särndal et al. [42, Result 3.7.2, p. 103].

As measures of the precision of the estimators $\hat{t}_d$ and $\hat{t}$, we consider their squared coefficients of variation, denoted by V_d and V, respectively:

$$V_d := \frac{\mathrm{Var}(\hat{t}_d)}{\big(\mathbb{E}[\hat{t}_d]\big)^2} \overset{[1]}{=} \frac{\mathrm{Var}(\hat{t}_d)}{t_d^2} \overset{(1.5)}{=} \frac{1}{t_d^2} \sum_{h=1}^{H_d} \Big(\frac{1}{n_{d,h}} - \frac{1}{N_{d,h}} \Big) N_{d,h}^2 S_{d,h}^2, \quad d \in \{1, \ldots, D\}, \tag{1.9a}$$

$$V := \frac{\mathrm{Var}(\hat{t})}{\big(\mathbb{E}[\hat{t}]\big)^2} \overset{[1]}{=} \frac{\mathrm{Var}(\hat{t})}{t^2} \overset{(1.8)}{=} \frac{1}{t^2} \sum_{d=1}^{D} \sum_{h=1}^{H_d} \Big(\frac{1}{n_{d,h}} - \frac{1}{N_{d,h}} \Big) N_{d,h}^2 S_{d,h}^2, \tag{1.9b}$$

where [1] follows from the unbiasedness of the estimators (1.4) and (1.7). The coefficient of variation is widely used as a consistent measure of estimation accuracy for totals and means of continuous variables (see, e.g., European Commission and Eurostat [13, Sec. 2.2, pp. 12-15]). The relative variances in (1.9) therefore provide standardized measures of estimator precision, facilitating meaningful comparisons across domains. Weighted variances have been extensively studied in the literature; see, for example, Dalenius [11], Yates [53], Hartley [17], Schaich and Münnich [38], or Choudhry, Hidiroglou and Rao [8].

The objective is to simultaneously minimize all domain-level relative variances V_d, $d \in \{1, \ldots, D\}$, as well as the overall relative variance V, subject to a fixed total sample size $n = \sum_{d=1}^{D} \sum_{h=1}^{H_d} n_{d,h}$, and the condition $n_{d,h} \leq N_{d,h}$ for all $h \in \{1, \ldots, H_d\}$ and $d \in \{1, \ldots, D\}$. To achieve this, we introduce pre-specified priority weights $\kappa_d > 0$, $d \in \{1, \ldots, D\}$, such that $\sum_{d=1}^{D} \kappa_d = 1$, and assume

$$V_d = \kappa_d T, \qquad d \in \{1, \ldots, D\}, \tag{1.10}$$

where T is an unknown nonnegative constant.

In this setting, each κ_d, $d \in \{1, \ldots, D\}$, specifies the proportion of T attributed to domain U_d. Accordingly, T serves as common baseline for the domain-level relative variances V_d, $d \in \{1, \ldots, D\}$, while the coefficients κ_d, $d \in \{1, \ldots, D\}$, determine how this baseline is distributed across domains. Since V_d measures the precision of the estimator $\hat{t}_d$, the κ_d, $d \in \{1, \ldots, D\}$, directly encode domain-specific precision priorities. This explicit control of domain-wise precisions through the parameters κ_d, $d \in \{1, \ldots, D\}$, motivates the term "controlled-precision" in the title of this thesis.

Under assumption (1.10), T governs not only the domain-level relative variances V_d, $d \in \{1, \ldots, D\}$, but also the overall relative variance V. To see the latter, note that (1.9) and (1.10) imply

$$V = \frac{1}{t^2} \Big(\sum_{d=1}^{D} t_d^2 \, \kappa_d \Big) \, T. \tag{1.11}$$

Hence, minimizing T simultaneously minimizes all V_d, $d \in \{1, \ldots, D\}$, and V.

We are now ready to precisely state the *multi-domain optimum sample allocation problem with controlled-precision under upper-bound constraints* (Problem 1.1.1), which constitutes the central focus of this thesis.

Problem 1.1.1. Let the following be given:

- strata population sizes $N_{d,h}$ and standard deviations $S_{d,h}$, $h \in \{1, \ldots, H_d\}$, $d \in \{1, \ldots, D\}$;
- domain totals t_d and priority weights $\kappa_d > 0$, $d \in \{1, \ldots, D\}$, satisfying $\sum_{d=1}^{D} \kappa_d = 1$;
- total sample size $n \in \left(0, \sum_{d=1}^{D} \sum_{h=1}^{H_d} N_{d,h}\right]$.

Determine an allocation vector $\boldsymbol{n} = \left(n_{d,h},\, h \in \{1, \ldots, H_d\},\, d \in \{1, \ldots, D\}\right)$ and a scalar T that solve the following optimization problem:

$$\underset{(T,\boldsymbol{n}) \in \mathbb{R}^{1+H}}{\text{minimize}} \quad T \tag{12}$$

$$\text{subject to} \quad \sum_{d=1}^{D} \sum_{h=1}^{H_d} n_{d,h} = n, \tag{13a}$$

$$\frac{1}{t_d^2 \kappa_d} \sum_{h=1}^{H_d} \left(\frac{1}{n_{d,h}} - \frac{1}{N_{d,h}} \right) N_{d,h}^2 S_{d,h}^2 = T, \qquad d \in \{1, \ldots, D\}, \tag{13b}$$

$$0 < n_{d,h} \leq N_{d,h}, \qquad h \in \{1, \ldots, H_d\}, \quad d \in \{1, \ldots, D\}, \tag{13c}$$

where $H = \sum_{d=1}^{D} H_d$ denotes the total number of strata.

We note the following regarding Problem 1.1.1:

- Constraints (13b) and (13c) together imply that any feasible T is nonnegative, thereby preserving assumption (1.10).
- In practice, the values of t_d and $S_{d,h}$ are rarely known. In such cases, a common approach is either to estimate them from previous studies (if available) or to use a proxy variable that is highly correlated with the study variable y and for which data are available. In the latter case, t_d and $S_{d,h}$ are computed from the proxy data.

This thesis is entirely devoted to the analysis and solution of Problem 1.1.1, including a detailed examination of its structure and the development of an algorithm, termed `RDCA`, to solve it. A comprehensive overview of the contributions is provided in Section 1.4, while Sections 1.2 and 1.3 review the related literature and present the motivation for studying this problem.

For a single domain (i.e., $D = 1$), so that the entire population belongs to that domain, Problem 1.1.1 reduces to the classical optimum sample allocation problem with upper-bound constraints, a fundamental problem in survey methodology. See Särndal et al. [42, Sec. 3.7.31, p. 104] and Särndal et al. [42, Rem. 12.7.1, p. 466] for discussion

and a sketch of the solution procedure, known as the *Recursive Neyman Algorithm* (`RNA`). A formal proof of the correctness of `RNA` is provided in Wesołowski, Wieczorkowski and Wójciak [46]. Alternative algorithms are discussed in Stenger and Gabler [41], Friedrich, Münnich, de Vries and Wagner [14], and Wright [49], with the latter two presenting methods that produce integer-valued solutions.

The `RNA` is of particular relevance to this thesis, as its central idea was used in the construction of the `RDCA` algorithm, one of the primary contributions of this work. Specifically, `RNA` computes tentative allocations according to the formula given in [46, Eq. (6), p. 1266] and then checks whether any allocation exceeds its stratum population size. If a violation occurs, the allocation for the corresponding stratum is fixed at its population size, and the allocations for the remaining strata are recomputed. The process continues iteratively until no violations occur. Although the optimality condition [see 46, Th. 1.1, p. 1266] is not verified directly in its explicit analytical form, the structural characterization of the solution implies that, once no feasibility violations remain, the resulting allocation necessarily satisfies all optimality conditions and is therefore optimal. This observation motivated the design of the `RDCA` algorithm.

1.2. Related Work

Choudhry et al. [8, Sec. 2.3, Sec. 4, pp. 24–25] studied a specific optimum allocation problem closely related to Problem 1.1.1, but in a somewhat different setting: estimators of domain means rather than domain totals were considered, the total sample size was minimized, and constraints corresponding to (13b) were formulated as inequalities. The authors solved the resulting problem using a general-purpose nonlinear programming (NLP) method based on the Newton-Raphson algorithm.

A notable general-purpose algorithm, applicable to a certain class of optimum allocation problems – including Problem 1.1.1 as a special case – was recently proposed by Willems [48]. Its correctness has been assessed through simulation studies.

A review of the existing literature suggests that, aside from general-purpose NLP numerical methods, no exact analytical algorithm currently exists for solving Problem 1.1.1. The classical monograph by Valliant, Dever and Kreuter [43] provides comprehensive coverage of nonlinear optimization methods for a wide range of optimum sample allocation problems, including problems closely related to Problem 1.1.1. In

particular, Chapter 5, *Mathematical Programming*, and Section 5.6, *Allocation for Domain Estimation*, of the monograph are especially relevant.

Niemiro and Wesołowski [32], Wesołowski and Wieczorkowski [45], Khan and Wesołowski [22], and Wesołowski [44], studied Problem 1.2.1, a simplified version of Problem 1.1.1 that omits the upper-bound constraints $n_{d,h} \leq N_{d,h}$ and includes an explicit positivity constraint on T.

Problem 1.2.1. Given the parameters specified in Problem 1.1.1, find an allocation vector $\boldsymbol{n} = \big(n_{d,h},\, h \in \{1, \ldots, H_d\},\, d \in \{1, \ldots, D\}\big)$ and a scalar T that solve the following optimization problem:

$$\underset{(T,\boldsymbol{n}) \in \mathbb{R}^{1+H}}{\text{minimize}} \quad T \tag{14}$$

$$\text{subject to} \quad \sum_{d=1}^{D} \sum_{h=1}^{H_d} n_{d,h} = n, \tag{15a}$$

$$\frac{1}{t_d^2 \kappa_d} \sum_{h=1}^{H_d} \left(\frac{1}{n_{d,h}} - \frac{1}{N_{d,h}} \right) N_{d,h}^2 S_{d,h}^2 = T, \qquad d \in \{1, \ldots, D\}, \tag{15b}$$

$$n_{d,h} > 0, \qquad h \in \{1, \ldots, H_d\}, \quad d \in \{1, \ldots, D\}, \tag{15c}$$

$$T > 0. \tag{15d}$$

Recall that in Problem 1.1.1, the non-negativity of the optimal T^* is ensured jointly by constraints (13b) and (13c). Since the upper-bound part of the constraints in (13c) is removed in Problem 1.2.1, an explicit constraint on T becomes necessary to preserve assumption (1.10), ensuring that the variances (1.9) remain nonnegative at the optimum.

The authors cited above derived closed-form analytical formulas for the optimal solution using the method of Lagrange multipliers. Their derivations led to an eigenvalue problem, in which the optimal solution corresponds to a particular eigenpair of the associated population matrix. Several of these results are used in this thesis and are recalled in detail in Appendix G, following the notation introduced in Chapter 2. The main limitation of this line of work is inherent in the problem formulation itself: because upper-bound constraints of the form $n_{d,h} \leq N_{d,h}$ are not imposed, the resulting allocations may exceed the available population sizes in some strata and thus be infeasible in practice.

1.3. Motivation

As presented in Section 1.1, Problem 1.1.1 constitutes the central focus of this thesis. The primary contribution of this work is the `RDCA` algorithm, which provides a solution to this problem. The motivation for this research is twofold. First, there is a practical motivation: this specific optimum allocation problem arises naturally in important applications, and obtaining an optimal solution is essential for both efficiency and precision. Second, there is a methodological motivation: to develop an analytical algorithm that produces an exact solution to this allocation problem. In this section, we elaborate on these motivations.

1.3.1. Motivation for studying the problem

As already indicated in Section 1.2, a simplified version of Problem 1.1.1, namely Problem 1.2.1, has been studied in the literature, and a closed-form expression for its solution is available. However, because this formulation does not include upper-bound constraints of the form $n_{d,h} \leq N_{d,h}$, the resulting allocations may exceed the available population sizes in certain strata, rendering them potentially infeasible in practice.

The motivation for studying Problem 1.1.1 is therefore clear: we seek an allocation that minimizes the variances (1.9) under a fixed total sample size, while ensuring feasibility with respect to strata population sizes and distributing an overall precision across domains according to the desired domain-wise priority weights.

The work undertaken in this thesis can therefore be seen as a natural continuation of the research initiated by Niemiro and Wesołowski [32] and subsequently extended by Wesołowski [44], which addressed Problem 1.2.1.

1.3.2. Motivation for an exact analytical algorithm – limitations of general-purpose NLP solvers

To the best of our knowledge, no exact analytical algorithm exists for solving Problem 1.1.1. Approximate solutions can be obtained using general-purpose nonlinear programming methods. These methods rely on iterative procedures to locate an extremum of the objective function and do not exploit the specific analytical structure of the allocation problem. Although NLP solvers often yield satisfactory numerical results

(mainly because approximate solutions are typically rounded to integer sample sizes in practice) they may suffer from numerical instability or convergence issues.

The `RDCA` algorithm proposed in this thesis is an analytical algorithm that produces an exact solution to Problem 1.1.1. It is based on the sufficient optimality conditions that characterize Problem 1.1.1 (formed in Theorem 3.2.8). These conditions provide an explicit formula for the optimal solution, which depends on a particular set of strata indices, referred to as the *optimal take-max* strata set. Once this set is identified, the allocation follows directly from the corresponding closed-form expressions. In this way, the algorithm leverages the problem's analytical structure to efficiently determine the optimal take-max strata set.

The proposed approach offers significant advantages in terms of numerical stability and interpretability compared to general-purpose NLP numerical methods. It avoids the convergence issues frequently encountered in iterative NLP solvers, which may fail when initialized from unsuitable starting points or, in some cases, may not converge at all. By exploiting the inherent analytical structure of the problem, the proposed algorithm combines theoretical transparency with practical robustness, particularly when the number of domains and strata is moderate.

1.3.3. Remark on allocations that are not necessarily integer-valued

In the formulation of Problem 1.1.1, we allow the solution to be non-integer, even though the actual sample sizes are (positive) integers. There are two main reasons for this choice.

First, introducing integer constraints significantly complicates the problem and typically requires additional combinatorial considerations. Therefore, we chose to address the non-integer version first, exploiting the analytical structure of the problem before tackling the more complex integer-constrained case.

Second, integer-valued allocation methods are typically much slower than their not-necessarily-integer counterparts. For example, in the classical optimum allocation problem with one-sided upper-bound constraints (i.e., Problem 1.1.1 with a single domain), the integer-valued `CapacityScaling` algorithm of Friedrich et al. [14] is significantly slower than the `RNA` algorithm [see 46, Sec. 4, p. 1270]. Computational efficiency is especially important when the number of strata is large (e.g., in the German

census application; Burgard and Münnich [7]) and becomes even more critical in iterative stratification procedures, where the number of iterations may reach millions; see, e.g., Baillargeon and Rivest [2], Barcaroli [3], Gunning and Horgan [16], Khan, Nand and Ahmad [21], Lednicki and Wieczorkowski [25].

A natural remedy for a not-necessarily-integer-valued allocation is to round the optimal non-integer solution produced by the algorithm. Overall, such a procedure may, in general, be considerably faster than integer-valued allocation algorithms. However, it should be emphasized that simple rounding can be problematic: there is no guarantee that it preserves optimality, and it may even lead to an infeasible solution, as noted by Friedrich et al. [14, Sec. 1, p. 3] in the context of the optimum allocation problem considered in that work.

Despite these drawbacks, empirical evidence suggests that the loss incurred by rounding may, in some cases, be negligible in practice. In particular, numerical experiments reported in Wesołowski, Wieczorkowski and Wójciak [47, Sec. 6, p. 497] show that, for Problem 1.1.1 with a single domain and strictly positive lower-bound constraints on stratum sample sizes, the values of the objective function obtained from the non-integer optimum allocation (both before and after rounding) and from the integer optimum allocation are practically indistinguishable. Moreover, infeasibility arising after rounding – specifically from violations of the total sample size constraint (13a) – can be avoided by employing the rounding method of Cont and Heidari [10], which preserves the integer sum of positive numbers.

Taking these considerations into account, it is justified to study Problem 1.1.1 in a non-integer formulation.

1.4. Main Results

The main contributions of this thesis concern the detailed analysis and algorithmic solution of the optimum allocation problem formulated as Problem 1.1.1.

The first major contribution is the theoretical analysis of Problem 1.1.1. We investigate its solvability (see Proposition 3.1.5) and establish sufficient optimality conditions (Theorem 3.2.8). All results are derived using rigorous mathematical reasoning and tools from convex optimization and matrix algebra. In particular, the

optimality conditions are obtained through the application of the Karush-Kuhn-Tucker framework and the analysis of an associated eigenvalue problem. These conditions are of central importance, as they provide the foundation not only for the design and correctness proofs of the algorithm developed in this thesis, but also for any potential future algorithms aimed at solving Problem 1.1.1.

The second major contribution is the *Recursive Domain Controlled Allocation* (`RDCA`) algorithm – an exact analytical algorithm that solves Problem 1.1.1. In addition to presenting the algorithm, this work rigorously establishes its definiteness and correctness (the latter using Hoare logic [19]). Furthermore, an implementation of the `RDCA` algorithm is provided in the R programming language, facilitating its application in real-world surveys.

Together, these results establish both the theoretical framework and the computational tools necessary for addressing Problem 1.1.1, contributing to a deeper understanding of its mathematical structure and to the development of efficient algorithms for its solution.

1.5. Organization of the Thesis

The dissertation is organized into seven chapters, followed by several appendices containing supplementary material.

Chapter 1 introduces the main topic of the thesis, formulating the problem from an application-oriented perspective, discussing the motivation, and stating the primary results. It also reviews the relevant literature and concludes with this structural overview of the dissertation.

Chapter 2 introduces an abstract model representing the population and total sample size. It presents basic properties of the model, which are essential for the proofs of correctness of the `RDCA` algorithm defined later in Chapter 4. The chapter also introduces several auxiliary functions that operate on domain-strata index sets, simplifying the notation in subsequent chapters.

Chapter 3 defines the optimum sample allocation problem studied in this thesis as an abstract optimization problem, referred to as the *Controlled-Precision Domain Allocation* (CPDA) problem. This definition builds upon the notation and the model introduced in Chapter 2. Furthermore, the chapter discusses the solvability of CPDA problem and derives sufficient optimality conditions using the Karush-Kuhn-Tucker framework.

These results characterize the structure of the optimal solution and provide the analytical foundation for the algorithmic approach developed in the subsequent chapters.

Chapter 4 presents the primary contribution of this work: the `RDCA` algorithm for solving the CPDA problem. After a formal definition, the chapter provides illustrative examples to clarify its operation. It also introduces a modified version of an existing algorithm, `DCA`, which serves as the base-case solver in the recursive structure of `RDCA`.

Chapter 5 is the first of two chapters dedicated to proving that the `RDCA` algorithm solves the CPDA problem, thereby establishing the so-called *total correctness* of the algorithm. This chapter introduces the notational conventions for program variables and proves that the algorithm is well-defined and guaranteed to terminate. Furthermore, it defines two key formal constructions – the Last-Branch Recursion Path and the take-max strata sets $\mathcal{V}_i$ – which are pivotal in establishing the algorithm's correctness.

Chapter 6 provides the formal proof of the total correctness of the `RDCA` algorithm. Building upon the structural groundwork laid in Chapter 5, this chapter systematically verifies that the algorithm's output satisfies the sufficient optimality conditions derived in Chapter 3. This is the most analytically intensive part of the thesis, reflecting the technical depth required to prove the algorithm's correctness.

Chapter 7 summarizes the primary results of this thesis and discusses their implications. It also outlines potential directions for future research, including performance improvements to the `RDCA` algorithm and extensions of the CPDA problem to incorporate additional constraints, such as integer allocations or lower bounds on sample sizes within strata.

In addition to the main chapters, several appendices are included. These provide supporting material on Perron-Frobenius theory, mathematical optimization, and topological background. The final appendix presents the multi-domain controlled-precision optimum sample allocation problem without upper-bound constraints, which serves as the basis for the problem considered in this thesis.

Some appendices, particularly those covering selected mathematical material (e.g., topological facts), are extensive and contain numerous detailed definitions, theorems, and related results. While most of them are familiar to mathematicians, they are included here to make the thesis self-contained and accessible to practitioners in survey sampling, who are also part of the intended readership.

Chapter 2

Population - Total Sample Size Model

In Section 1.1, we formulated Problem 1.1.1, which is the central focus of this thesis. This problem depends on numerous parameters, such as strata sizes, standard deviations of the study variable within strata, the total sample size, and other quantities.

To streamline notation and facilitate subsequent analysis, this chapter introduces an abstract model representing the population and sampling quantities, which constitute the parameters of Problem 1.1.1. We define the model as a quintuple whose elements satisfy specific conditions. Additionally, the chapter presents several auxiliary set functions defined on the set of domain-stratum indices specified by the model. These functions provide a consistent framework for referencing domains and strata throughout the remainder of the thesis.

We begin by defining notational conventions for points indexed by pairs of natural numbers.

2.1. Points Indexed by $\mathbb{N}^2$

Definition 2.1.1 (Lexicographic order on $\mathbb{N}^2$)**.** Let $i = (i_1, i_2) \in \mathbb{N}^2$ and $j = (j_1, j_2) \in \mathbb{N}^2$. We say that i is less than j in lexicographic order, written $i <_{\text{lex}} j$, if

$$(i_1 < j_1) \quad \text{or} \quad (i_1 = j_1 \ \text{ and } \ i_2 < j_2). \tag{2.1}$$

Example 2.1.1 (Lexicographic order on $\mathbb{N}^2$)**.** For instance:

$$(2, 3) <_{\text{lex}} (3, 2),$$
$$(2, 3) <_{\text{lex}} (2, 5).$$

Let $I \subset \mathbb{N}^2$ be nonempty and finite. Suppose $a_i \in \mathbb{R}$ for all $i \in I$, and $b \in \mathbb{R}$. We write:

$$\begin{aligned} (a_i,\, i \in I) &:= (a_{i_1},\, \ldots,\, a_{i_{|I|}}), \\ \big(b,\, (a_i,\, i \in I)\big) &:= (b,\, a_{i_1},\, \ldots,\, a_{i_{|I|}}), \end{aligned} \tag{2.2}$$

where $i_k \in I$, $k \in \{1,\, \ldots,\, |I|\}$, and $i_1 <_{\text{lex}} \cdots <_{\text{lex}} i_{|I|}$.

For example, let $I = \{(2,3),\, (1,3),\, (4,8),\, (2,5)\}$. Then, for $a_{d,h} \in \mathbb{R}$, $(d,h) \in I$, and $b \in \mathbb{R}$, we have

$$\big(b, (a_{d,h},\, (d,h) \in I)\big) = (b,\, a_{1,3},\, a_{2,3},\, a_{2,5},\, a_{4,8}). \tag{2.3}$$

2.2. Model Definition

To provide a formal and compact notation for the population and sampling quantities appearing as parameters in Problem 1.1.1, we introduce the notion of a *model*. A model abstracts these quantities into a single object, collecting all numerical requirements so that they are explicitly specified and can be consistently referenced wherever needed. Moreover, this allows Problem 1.1.1 to be expressed in a purely abstract form as Problem 3.1.1 (presented later in Chapter 3), where the model p encapsulates all relevant parameters.

Definition 2.2.1. A *model* is defined as a quintuple $(\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n)$, where:

$$\mathcal{H} \subset \mathbb{N}^2, \quad 0 < |\mathcal{H}| < \infty, \tag{2.4a}$$

$$\boldsymbol{N} = (N_{d,h},\ (d,h) \in \mathcal{H}) \in \mathbb{N}^{|\mathcal{H}|}, \tag{2.4b}$$

$$\boldsymbol{S} = (S_{d,h},\ (d,h) \in \mathcal{H}) \in \mathbb{R}_+^{|\mathcal{H}|}, \tag{2.4c}$$

$$\boldsymbol{\rho} = (\rho_d,\, d \in \mathcal{D}) \in \mathbb{R}_+^{|\mathcal{D}|}, \tag{2.4d}$$

$$n \in \Big(0,\, \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big], \tag{2.4e}$$

with $\mathcal{D} := \{d \in \mathbb{N}\colon\ \exists h \in \mathbb{N},\, (d,h) \in \mathcal{H}\}$.

Definition 2.2.2. The class $\mathcal{P}$ of models is defined by

$$\mathcal{P} := \big\{(\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n)\colon\ (2.4) \text{ holds}\big\}. \tag{2.5}$$

By Definitions 2.2.1 and 2.2.2, an element $p \in \mathcal{P}$ is called a model. The components of a model $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n)$ represent the parameters of Problem 1.1.1 as follows.

The set $\mathcal{H}$ represents all domain-stratum index pairs in the population; that is, $(d, h) \in \mathcal{H}$ indexes stratum $U_{d,h}$, with d indexing its domain U_d. For simplicity, we often refer to a domain U_d by d, and to a stratum $U_{d,h}$ by (d, h) (or sometimes by h if d is clear).

In Section 1.1 (see (1.1) and (1.3)), domains and strata were assumed to be indexed by consecutive natural numbers starting from 1, i.e., $d \in \{1, \ldots, D\}$ and $h \in \{1, \ldots, H_d\}$. Definition 2.2.1 does not impose this restriction on $\mathcal{H}$, which entails no conceptual change; it is merely a matter of labelling.

The vector $\boldsymbol{N}$ represents the sizes of the strata, $\boldsymbol{S}$ the standard deviations of the study variable within strata, and n the total sample size.

The vector $\boldsymbol{\rho}$ represents the product of the domain total t_d and the square root of the priority weight κ_d, i.e., $\rho_d = t_d\sqrt{\kappa_d}$, $d \in \mathcal{D}$ (with $\mathcal{D}$ as in Def. 2.2.1). The assumption $\sum_{d=1}^{D} \kappa_d = 1$, stated in Problem 1.1.1, does not affect the solution and is therefore not included in the model definition.

Example 2.2.1 (A model)**.** Consider a population with two domains and five strata (2+3), having domain totals $t_1 = 2$, $t_2 = 3$, priority weights $\kappa_1 = 0.4$, $\kappa_2 = 0.6$, and a total sample size $n = 350$. The corresponding model $(\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$ is given by

$$\begin{aligned}
\mathcal{H} &= \{(1,\,1),\,(1,\,2),\,(2,\,1),\,(2,\,2),\,(2,\,3)\},\\
\boldsymbol{N} &= (N_{1,1},\,N_{1,2},\,N_{2,1},\,N_{2,2},\,N_{2,3}) = (100,\,200,\,150,\,40,\,50),\\
\boldsymbol{S} &= (S_{1,1},\,S_{1,2},\,S_{2,1},\,S_{2,2},\,S_{2,3}) = (10.5,\,2,\,50.2,\,30.7,\,20),\\
\boldsymbol{\rho} &= (\rho_1,\,\rho_2) = (1.26,\,2.32),\\
n &= 350.
\end{aligned} \tag{2.6}$$

The strata in the first domain are $\{(1,1),\,(1,2)\}$, and in the second domain $\{(2,1),\,(2,2),\,(2,3)\}$.

Lemma 2.2.1. *Let* $(\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$, $\mathcal{A} \subseteq \mathcal{H}$, *and* $\boldsymbol{z} = (z_{d,h},\,(d,h) \in \mathcal{H}) \in \mathbb{R}_+^{|\mathcal{H}|}$ *be such that*

$$z_{d,h} \geq N_{d,h}, \qquad (d,h) \in \mathcal{A}, \tag{2.7a}$$

$$n = \sum_{(d,h)\in\mathcal{H}} z_{d,h}. \tag{2.7b}$$

Then

$$\Big(n > \sum_{(d,h)\in\mathcal{A}} N_{d,h}\Big) \quad \vee \quad \Big(\mathcal{A} = \mathcal{H} \quad \wedge \quad n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big). \tag{2.8}$$

Proof. By (2.7b), we have

$$n = \sum_{(d,h)\in\mathcal{H}} z_{d,h} = \sum_{(d,h)\in\mathcal{A}} z_{d,h} + \sum_{(d,h)\in\mathcal{H}\setminus\mathcal{A}} z_{d,h} \overset{(2.7a)}{\geq} \sum_{(d,h)\in\mathcal{A}} N_{d,h} + \sum_{(d,h)\in\mathcal{H}\setminus\mathcal{A}} z_{d,h}. \tag{2.9}$$

Since $z_{d,h} > 0$ for all $(d,h) \in \mathcal{H}$, the last sum in (2.9) is strictly positive if $\mathcal{A} \subsetneq \mathcal{H}$, and thus

$$n > \sum_{(d,h)\in\mathcal{A}} N_{d,h}. \tag{2.10}$$

In the remaining case $\mathcal{A} = \mathcal{H}$, we obtain

$$n \overset{(2.9)}{\geq} \sum_{(d,h)\in\mathcal{H}} N_{d,h} \overset{(2.4e)}{\geq} n, \tag{2.11}$$

and therefore $n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}$. □

Definition 2.2.3 (Family of admissible sets)**.** Let $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$. The *family of admissible sets* for p is defined by

$$\mathcal{F}_p := \Big\{\mathcal{A} \subset \mathcal{H}\colon\ n > \sum_{(d,h)\in\mathcal{A}} N_{d,h}\Big\}. \tag{2.12}$$

Remark 2.2.1. Let $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$. Then, by the condition (2.4e),

$$\mathcal{H} \notin \mathcal{F}_p, \tag{2.13}$$

that is, every admissible set $\mathcal{A} \in \mathcal{F}_p$ is a proper subset of $\mathcal{H}$.

2.3. Auxiliary Functions

In this section, we define several set functions that will be used extensively throughout the thesis to refer to domains and strata in a given model $p \in \mathcal{P}$.

Definition 2.3.1 (Set functions on domain-stratum index sets)**.** Let $(\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$. For any $\mathcal{A} \subseteq \mathcal{H}$, we define the following set functions:

$$\delta(\mathcal{A}) := \{d \in \mathbb{N}\colon\ \exists h \in \mathbb{N},\ (d,h) \in \mathcal{A}\}, \tag{2.14a}$$

$$\eta_d(\mathcal{A}) := \{h \in \mathbb{N}\colon\ (d,h) \in \mathcal{A}\}, \qquad d \in \delta(\mathcal{H}), \tag{2.14b}$$

$$\gamma_{\mathcal{O}}(\mathcal{A}) := \{(d,h) \in \mathcal{A}\colon\ d \in \mathcal{O}\}, \qquad \mathcal{O} \subseteq \delta(\mathcal{H}), \tag{2.14c}$$

$$\gamma_d(\mathcal{A}) := \gamma_{\{d\}}(\mathcal{A}), \qquad d \in \delta(\mathcal{H}). \tag{2.14d}$$

Definition 2.3.2. For any $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$, we define the following functions:

$$A_{d,h}(p) := \frac{N_{d,h} S_{d,h}}{\rho_d}, \qquad (d,h) \in \mathcal{H}, \tag{2.15a}$$

$$c_d(p) := \sum_{h \in \eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{N_{d,h}}, \qquad d \in \delta(\mathcal{H}). \tag{2.15b}$$

According to Definition 2.3.1, for a given $\mathcal{A} \subseteq \mathcal{H}$, $\delta(\mathcal{A})$ is the set of domain indices corresponding to the strata in $\mathcal{A}$, i.e., the projection of $\mathcal{A}$ onto its first coordinate. Similarly, for any $d \in \delta(\mathcal{H})$, $\eta_d(\mathcal{A})$ is the set of strata indices h within domain d such that $(d,h) \in \mathcal{A}$, i.e., it collects the second coordinates of the pairs in $\mathcal{A}$ with first coordinate d.

Example 2.3.1. Let $(\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$ be such that $\mathcal{H} = \{(d,h) \in \mathbb{N}\colon\ d \le 5,\ h \le 4\}$ and let $\mathcal{A} = \{(2,\,1),\,(2,\,3),\,(3,\,4),\,(5,\,2)\}$. Then:

$$\begin{aligned}
\delta(\emptyset) &= \emptyset, & \gamma_2(\mathcal{A}) &= \{(2,\,1),\,(2,\,3)\},\\
\delta(\mathcal{H}) &= \{1,\,2,\,3,\,4,\,5\}, & \gamma_4(\mathcal{A}) &= \emptyset,\\
\delta(\mathcal{A}) &= \{2,\,3,\,5\}, & \gamma_{\{2,3\}}(\mathcal{A}) &= \{(2,\,1),\,(2,\,3),\,(3,\,4)\},\\
\eta_2(\mathcal{A}) &= \{1,\,3\}, & \gamma_9(\mathcal{A}) &\text{ – not defined as } 9 \notin \delta(\mathcal{H}).
\end{aligned} \tag{2.16}$$

Corollary 2.3.1. *Let $(\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$. For any $\mathcal{A},\, \mathcal{B} \subseteq \mathcal{H}$ and any $\mathcal{O},\, \mathcal{T} \subseteq \delta(\mathcal{H})$, the following identities hold:*

$$\gamma_{\mathcal{O}}(\mathcal{A} \cup \mathcal{B}) = \gamma_{\mathcal{O}}(\mathcal{A}) \cup \gamma_{\mathcal{O}}(\mathcal{B}), \tag{2.17a}$$

$$\gamma_{\mathcal{O}}(\mathcal{A} \cap \mathcal{B}) = \gamma_{\mathcal{O}}(\mathcal{A}) \cap \gamma_{\mathcal{O}}(\mathcal{B}), \tag{2.17b}$$

$$\gamma_{\mathcal{O}}(\mathcal{A} \setminus \mathcal{B}) = \gamma_{\mathcal{O}}(\mathcal{A}) \setminus \gamma_{\mathcal{O}}(\mathcal{B}) = \gamma_{\mathcal{O}}(\mathcal{A}) \setminus \mathcal{B}, \tag{2.17c}$$

$$\gamma_{\mathcal{O} \cup \mathcal{T}}(\mathcal{A}) = \gamma_{\mathcal{O}}(\mathcal{A}) \cup \gamma_{\mathcal{T}}(\mathcal{A}), \tag{2.17d}$$

$$\gamma_{\mathcal{O} \cap \mathcal{T}}(\mathcal{A}) = \gamma_{\mathcal{O}}(\mathcal{A}) \cap \gamma_{\mathcal{T}}(\mathcal{A}). \tag{2.17e}$$

Moreover,

$$\delta(\mathcal{A} \setminus \mathcal{B}) = \delta(\mathcal{A}) \setminus \{d \in \delta(\mathcal{A})\colon\ \mathcal{B} \supseteq \gamma_d(\mathcal{A})\}. \tag{2.17f}$$

Proof. Most of these relations are straightforward. In the following, we prove only (2.17c) and (2.17f).

Let $\mathcal{A},\, \mathcal{B} \subseteq \mathcal{H}$ and $\mathcal{O} \subseteq \delta(\mathcal{H})$.

(2.17c): Let $\mathcal{B}^{\complement} := \mathcal{H} \setminus \mathcal{B}$. Since $\gamma_{\mathcal{O}}(\mathcal{B} \cup \mathcal{B}^{\complement}) = \gamma_{\mathcal{O}}(\mathcal{H})$, by (2.17a),

$$\gamma_{\mathcal{O}}(\mathcal{B}^{\complement}) = \gamma_{\mathcal{O}}(\mathcal{H}) \setminus \gamma_{\mathcal{O}}(\mathcal{B}). \tag{2.18}$$

Hence,

$$\begin{aligned} \gamma_{\mathcal{O}}(\mathcal{A} \setminus \mathcal{B}) &\overset{(2.17b)}{=} \gamma_{\mathcal{O}}(\mathcal{A}) \cap \gamma_{\mathcal{O}}(\mathcal{B}^{\complement}) \\ &\overset{(2.18)}{=} \gamma_{\mathcal{O}}(\mathcal{A}) \cap \big(\gamma_{\mathcal{O}}(\mathcal{H}) \setminus \gamma_{\mathcal{O}}(\mathcal{B})\big) \\ &\overset{[1]}{=} \gamma_{\mathcal{O}}(\mathcal{A}) \setminus \gamma_{\mathcal{O}}(\mathcal{B}) = \gamma_{\mathcal{O}}(\mathcal{A}) \setminus \mathcal{B}, \end{aligned} \tag{2.19}$$

where [1] holds because $\gamma_{\mathcal{O}}(\mathcal{A}) \subseteq \gamma_{\mathcal{O}}(\mathcal{H})$.

(2.17f): For any $d \in \delta(\mathcal{A})$,

$$\begin{aligned} d \in \delta(\mathcal{A} \setminus \mathcal{B}) \quad &\Longleftrightarrow \quad \gamma_d(\mathcal{A} \setminus \mathcal{B}) \neq \emptyset \\ &\overset{(2.17c)}{\Longleftrightarrow} \quad \gamma_d(\mathcal{A}) \setminus \mathcal{B} \neq \emptyset \\ &\Longleftrightarrow \quad \gamma_d(\mathcal{A}) \nsubseteq \mathcal{B}. \end{aligned} \tag{2.20}$$

Therefore,

$$\delta(\mathcal{A} \setminus \mathcal{B}) = \{d \in \delta(\mathcal{A}) \colon\ \gamma_d(\mathcal{A}) \nsubseteq \mathcal{B}\} = \delta(\mathcal{A}) \setminus \{d \in \delta(\mathcal{A}) \colon\ \gamma_d(\mathcal{A}) \subseteq \mathcal{B}\}. \tag{2.21}$$

□

Remark 2.3.1 (Blocked domains)**.** Let $(\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$ and $\mathcal{B} \subseteq \mathcal{H}$. By (2.17f), for any $d \in \delta(\mathcal{H})$, the following equivalence holds:

$$d \in \delta(\mathcal{H} \setminus \mathcal{B}) \quad \Longleftrightarrow \quad \gamma_d(\mathcal{H}) \nsubseteq \mathcal{B}. \tag{2.22}$$

If the set $\mathcal{B}$ is interpreted as the collection of *blocked* strata, then a domain $d \in \delta(\mathcal{H})$ is said to be *blocked* with respect to $\mathcal{B}$ if and only if all of its strata belong to $\mathcal{B}$ (i.e., $\gamma_d(\mathcal{H}) \subseteq \mathcal{B}$). Consequently, by (2.22), $\delta(\mathcal{H} \setminus \mathcal{B})$ represents the set of domains that remain *unblocked* with respect to $\mathcal{B}$.

Although the term "blocked" is used in Remark 2.3.1 in an abstract set-theoretic sense, it is given a concrete operational meaning later in the context of the `RDCA` algorithm. This notion and its interpretation play a central role in Section 6.2 of Chapter 6, where the correctness of the algorithm is established.

Corollary 2.3.2. *Let* $(\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$, $\mathcal{A} \subseteq \mathcal{H}$, *and* $\mathcal{O} \subseteq \delta(\mathcal{H})$. *For any* $(d,h) \in \mathcal{H}$,

$$\Big(d \in \delta(\mathcal{H} \setminus \mathcal{A}) \setminus \mathcal{O} \quad \wedge \quad h \in \eta_d(\mathcal{H} \setminus \mathcal{A})\Big) \quad \Longleftrightarrow \quad (d,h) \in \gamma_{\delta(\mathcal{H})\setminus\mathcal{O}}(\mathcal{H}) \setminus \mathcal{A}. \tag{2.23}$$

Proof. Let $\mathcal{B} := \mathcal{H} \setminus \mathcal{A}$, and fix an arbitrary $(d,h) \in \mathcal{H}$.

By Definition 2.3.1 of the functions η_d, δ, and $\gamma_{\delta(\mathcal{H})\setminus\mathcal{O}}$, we have:

$$h \in \eta_d(\mathcal{B}) \quad \Longleftrightarrow \quad (d,h) \in \mathcal{B}, \tag{2.24a}$$

$$(d,h) \in \mathcal{B} \quad \Longrightarrow \quad d \in \delta(\mathcal{B}), \tag{2.24b}$$

$$\Big(d \in \delta(\mathcal{H}) \setminus \mathcal{O} \quad \wedge \quad (d,h) \in \mathcal{B}\Big) \quad \Longleftrightarrow \quad (d,h) \in \gamma_{\delta(\mathcal{H})\setminus\mathcal{O}}(\mathcal{B}). \tag{2.24c}$$

Combining these, we obtain

$$\begin{aligned}
&\Big(d \in \delta(\mathcal{B}) \setminus \mathcal{O} \quad \wedge \quad h \in \eta_d(\mathcal{B})\Big) \\
&\quad \overset{(2.24a)}{\Longleftrightarrow} \quad \Big(d \in \delta(\mathcal{B}) \quad \wedge \quad d \in \delta(\mathcal{H}) \setminus \mathcal{O} \quad \wedge \quad (d,h) \in \mathcal{B}\Big) \\
&\quad \overset{[1]}{\Longleftrightarrow} \quad \Big(d \in \delta(\mathcal{H}) \setminus \mathcal{O} \quad \wedge \quad (d,h) \in \mathcal{B}\Big) \\
&\quad \overset{(2.24c)}{\Longleftrightarrow} \quad (d,h) \in \gamma_{\delta(\mathcal{H})\setminus\mathcal{O}}(\mathcal{B}) \overset{(2.17c)}{=} \gamma_{\delta(\mathcal{H})\setminus\mathcal{O}}(\mathcal{H}) \setminus \mathcal{A},
\end{aligned} \tag{2.25}$$

where [1] holds because the first conjunct is implied by the third, see (2.24b). □

2.4. Sums and Cartesian Products over the Set of Domain-Stratum Indices $\mathcal{H}$

In many situations, it is necessary to reorganize sums over the set of domain-stratum indices $\mathcal{H}$ as iterated sums over domains and strata. To illustrate, consider a model $(\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$ and let $\mathcal{A} \subseteq \mathcal{H}$. Using the set functions defined in Definition 2.3.1, a sum of $a_{d,h} \in \mathbb{R}$ over $(d,h) \in \mathcal{A}$ can be written as

$$\sum_{(d,h)\in\mathcal{A}} a_{d,h} = \sum_{d\in\delta(\mathcal{A})} \sum_{h\in\eta_d(\mathcal{A})} a_{d,h}. \tag{2.26}$$

Given a nonempty set $\mathcal{A} \subseteq \mathcal{H}$ and a family of sets $\{X_{d,h}\}_{(d,h)\in\mathcal{A}}$, the Cartesian product over $\mathcal{A}$ is defined by

$$\bigtimes_{(d,h)\in\mathcal{A}} X_{d,h} := \big\{(a_{d,h},\, (d,h) \in \mathcal{A})\colon\ a_{d,h} \in X_{d,h} \text{ for all } (d,h) \in \mathcal{A}\big\}. \tag{2.27}$$

Chapter 3

The Controlled-Precision Domain Allocation Problem

In this chapter, we rewrite the optimum allocation Problem 1.1.1 at a fully abstract level, using the notational conventions and definitions introduced in Chapter 2. The resulting formulation, stated as Problem 3.1.1, is referred to as the *Controlled-Precision Domain Allocation*, or CPDA, problem.

The purpose of this rewrite is to express Problem 1.1.1 entirely in terms of the population-total sample size model and the associated functions defined in Chapter 2. This allows for a compact and uniform notation that simplifies both the statement of results and the presentation of their proofs, while also providing a more formal and rigorous framework.

Within this abstract setting, in Section 3.1, we establish the solvability of the CPDA problem. This is done in two steps: first, we show that the feasible set is nonempty; second, we prove that the objective function attains its minimum over this set.

In Section 3.2, we derive sufficient optimality conditions for a solution to the CPDA problem. These conditions form the theoretical foundation for the algorithmic developments presented in Chapter 4.

3.1. Problem Definition and Solvability

Throughout this section, for any subset $X \subseteq \mathbb{R}^n$ ($n \in \mathbb{N}$), whenever we refer to X as a metric space, we mean the metric space (X, m), where m is the standard Euclidean metric restricted to X. We omit explicit reference to m in subsequent analysis as it is not required for the results.

Problem 3.1.1 (CPDA)**.** For a given model $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$, the CPDA($p$) optimization problem is defined as:

$$\underset{(T,\,\boldsymbol{x})\,\in\,\mathbb{R}\times\mathbb{R}_+^{|\mathcal{H}|}}{\text{minimize}} \quad T \tag{3.1}$$

$$\text{subject to} \quad \sum_{(d,h)\in\mathcal{H}} x_{d,h} - n = 0, \tag{3.2a}$$

$$\sum_{h\in\eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) - T = 0, \qquad d \in \delta(\mathcal{H}), \tag{3.2b}$$

$$x_{d,h} - N_{d,h} \le 0, \qquad (d,h) \in \mathcal{H}, \tag{3.2c}$$

where $(T,\, \boldsymbol{x}) = \big(T,\, (x_{d,h},\, (d,h) \in \mathcal{H})\big)$ is the optimization variable, and the functions δ, η_d, $A_{d,h}$, and c_d are as defined in Definitions 2.3.1 and 2.3.2 in the context of p.

The formulation of Problem 3.1.1 uses the symbol $\boldsymbol{x}$ (instead of $\boldsymbol{n}$ as in Problem 1.1.1) to emphasize that the solution to CPDA is not required to be integer-valued.

Corollary 3.1.1. *Let $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$, and let*

$$F := \{(T,\, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}\colon\ (3.2) \text{ holds for } p\} \tag{3.3}$$

be the feasible set of the CPDA*(p) problem. Then, for any $(T,\, \boldsymbol{x}) \in F$, the following hold:*

$$T \ge 0, \tag{3.4a}$$

$$T = 0 \quad \Longleftrightarrow \quad \boldsymbol{x} = \boldsymbol{N}. \tag{3.4b}$$

Proof. The result follows directly from constraints (3.2b) and (3.2c). □

Lemma 3.1.2. *Let $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$. Suppose that $\eta_d(\mathcal{H}) = \{1,\, \ldots,\, |\eta_d(\mathcal{H})|\}$ for all $d \in \delta(\mathcal{H})$, and that $|\eta_i(\mathcal{H})| \ge 2$ for some $i \in \delta(\mathcal{H})$.*

Let $\widetilde{\mathcal{H}} := \{(d,h) \in \mathcal{H}\colon\ h \ne 1\}$ and define the set K by

$$K := \{(T,\, \widetilde{\boldsymbol{x}}) \in \mathbb{R} \times \mathbb{R}_+^{|\widetilde{\mathcal{H}}|}\colon\ (3.6) \text{ holds}\}, \tag{3.5}$$

where:

$$T \ge \max_{d\in\delta(\mathcal{H})} \Big(\sum_{h\in\eta_d(\widetilde{\mathcal{H}})} \tfrac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) + \tfrac{[A_{d,1}(p)]^2}{N_{d,1}} \Big), \tag{3.6a}$$

$$\widetilde{\boldsymbol{x}} = (x_{d,h},\, (d,h) \in \widetilde{\mathcal{H}}) \in \bigtimes_{(d,h)\in\widetilde{\mathcal{H}}} (0,\, N_{d,h}]. \tag{3.6b}$$

Then $K \ne \emptyset$ and K is a connected metric space.

Proof. Observe that $\widetilde{\mathcal{H}} \neq \emptyset$, since there exists $i \in \delta(\mathcal{H})$ such that $|\eta_i(\mathcal{H})| \geq 2$.

To show that $K \neq \emptyset$, choose $\widetilde{\boldsymbol{x}} = \big(\frac{N_{d,h}}{2},\ (d,h) \in \widetilde{\mathcal{H}}\big)$ and let $T \in \mathbb{R}$ be sufficiently large so that (3.6a) is satisfied. Then $(T,\ \widetilde{\boldsymbol{x}}) \in K$, and hence $K \neq \emptyset$.

To prove connectedness, let $X := \bigtimes_{(d,h)\in\widetilde{\mathcal{H}}} (0,\ N_{d,h}]$, and define $T_{min}\colon X \to \mathbb{R}$ by

$$T_{min}(\widetilde{\boldsymbol{x}}) := \max_{d\in\delta(\mathcal{H})} \Big(\sum_{h\in\eta_d(\widetilde{\mathcal{H}})} \tfrac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) + \tfrac{[A_{d,1}(p)]^2}{N_{d,1}} \Big). \tag{3.7}$$

For each $(d,h) \in \widetilde{\mathcal{H}}$, the function $k_{d,h}\colon X \to \mathbb{R}$, $k_{d,h}(\widetilde{\boldsymbol{x}}) := \frac{1}{x_{d,h}}$, is convex. Moreover, by convention, a sum over an empty set is taken to be 0, which is convex. Consequently, for each $d \in \delta(\mathcal{H})$, the function inside the maximum is convex on X; hence, $T_{\min}$, being the pointwise maximum of finitely many convex functions, is convex. (see Prop. D.1). Therefore, its epigraph

$$\operatorname{epi} T_{min} := \big\{(T,\ \widetilde{\boldsymbol{x}}) \in \mathbb{R} \times X\colon\ T \geq T_{min}(\widetilde{\boldsymbol{x}})\big\} \tag{3.8}$$

is a convex set. Since $K = \operatorname{epi} T_{\min}$, Theorem E.3.2 implies that K is a connected metric space. □

Lemma 3.1.3. *Let $p = (\mathcal{H},\ \boldsymbol{N},\ \boldsymbol{S},\ \boldsymbol{\rho},\ n) \in \mathcal{P}$ be a model satisfying the assumptions of Lemma 3.1.2, and let the sets $\widetilde{\mathcal{H}}$ and K be defined as in that lemma.*

Define the function $k\colon K \to \mathbb{R}$ by

$$k(T,\ \widetilde{\boldsymbol{x}}) := \sum_{d\in\delta(\mathcal{H})} \frac{[A_{d,1}(p)]^2}{T - \sum_{h\in\eta_d(\widetilde{\mathcal{H}})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} + c_d(p)} + \sum_{(d,h)\in\widetilde{\mathcal{H}}} x_{d,h}, \tag{3.9}$$

where $\widetilde{\boldsymbol{x}} = (x_{d,h},\ (d,h) \in \widetilde{\mathcal{H}})$.

Then

$$k(K) = \Big(0,\ \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big]. \tag{3.10}$$

Proof. First, recall that $\widetilde{\mathcal{H}} \neq \emptyset$ by the assumptions of Lemma 3.1.2. Furthermore, the constraint (3.6a) in the definition of K is equivalent to

$$T - \sum_{h\in\eta_d(\widetilde{\mathcal{H}})} \tfrac{[A_{d,h}(p)]^2}{x_{d,h}} \geq -c_d(p) + \tfrac{[A_{d,1}(p)]^2}{N_{d,1}}, \qquad d \in \delta(\mathcal{H}), \tag{3.11}$$

which ensures that each denominator in (3.9) is nonzero. Thus, k is well-defined.

To prove the lemma, we show that $\inf_K k = 0 \notin k(K)$ and $\max_K k = \sum_{(d,h)\in\mathcal{H}} N_{d,h}$. From this, together with the continuity of k and the connectedness of K, the result follows.

$\underline{\inf_K k = 0 \notin k(K).}$

- $k(K) \subset (0, +\infty)$.

  Indeed, the first sum in (3.9) is strictly positive due to constraint (3.6a) (equivalently, (3.11)), while the second sum is strictly positive due to (3.6b).

- $\forall \epsilon > 0,\ k^{-1}((0, \epsilon]) \neq \emptyset$.

  Choosing $x_{d,h} \in (0, N_{d,h}]$ sufficiently small for all $(d, h) \in \widetilde{\mathcal{H}}$ and $T \in \mathbb{R}$ large enough (consistent with (3.6a)) ensures the preimage is nonempty.

Hence, $\inf_K k = 0$. Since $k(K) \subset (0, +\infty)$, we also have $0 \notin k(K)$.

$\underline{\max_K k = \sum_{(d,h)\in\mathcal{H}} N_{d,h}.}$

Let

$$(T_0,\, \widetilde{\boldsymbol{x}}_0) := \big(0,\, (N_{d,h},\, (d,h) \in \widetilde{\mathcal{H}})\big) \in K. \tag{3.12}$$

We demonstrate that $(T_0,\, \widetilde{\boldsymbol{x}}_0) \in \arg\max_K k$, by observing that $(T_0,\, \widetilde{\boldsymbol{x}}_0)$ simultaneously maximizes both summands in the definition of k.

- Regarding the first summand, observe that for each $d \in \delta(\mathcal{H})$,

  $$\underset{(T,\widetilde{\boldsymbol{x}})\in K}{\arg\max} \frac{[A_{d,1}(p)]^2}{T - \sum_{h\in\eta_d(\widetilde{\mathcal{H}})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} + c_d(p)} = \underset{(T,\widetilde{\boldsymbol{x}})\in K}{\arg\min}\Big(T - \sum_{h\in\eta_d(\widetilde{\mathcal{H}})} \frac{[A_{d,h}(p)]^2}{x_{d,h}}\Big). \tag{3.13}$$

  Moreover, the system of inequalities in (3.11), which is part of the definition of K, holds with equality for all $d \in \delta(\mathcal{H})$ at $(T,\, \widetilde{\boldsymbol{x}}) = (T_0,\, \widetilde{\boldsymbol{x}}_0)$.

  Therefore, by (3.13), the point $(T_0,\, \widetilde{\boldsymbol{x}}_0)$ simultaneously maximizes each term of the first summand and, consequently, the entire first summand.

- For any $(T,\, \widetilde{\boldsymbol{x}}) \in K$, by (3.6b), we have $x_{d,h} \leq N_{d,h}$ for all $(d, h) \in \widetilde{\mathcal{H}}$. Since $\widetilde{\boldsymbol{x}}_0 = (N_{d,h},\, (d,h) \in \widetilde{\mathcal{H}})$, it follows that

  $$(T_0,\, \widetilde{\boldsymbol{x}}_0) \in \underset{(T,\widetilde{\boldsymbol{x}})\in K}{\arg\max} \sum_{(d,h)\in\widetilde{\mathcal{H}}} x_{d,h}. \tag{3.14}$$

Because each of the two summands in the definition of k is maximized at $(T_0,\, \widetilde{\boldsymbol{x}}_0)$, the function k attains its maximum at that point, namely $k(T_0,\, \widetilde{\boldsymbol{x}}_0) = \sum_{(d,h)\in\mathcal{H}} N_{d,h}$.

Since K is a connected metric space (Lem. 3.1.2) and the mapping $k\colon K \to \mathbb{R}$ is continuous, Theorem E.4.3 implies that the image $k(K)$ is connected in metric space $\mathbb{R}$. Consequently, Theorem E.3.1 implies that the image $k(K)$ is an interval. Together with the infimum and maximum bounds established above, we conclude

$$k(K) = \Big(0,\, \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big]. \tag{3.15}$$

□

Proposition 3.1.4. *Let* $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$*, and let*

$$F := \{(T,\, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}\colon\ (3.2) \text{ holds for } p\}, \tag{3.16}$$

be the feasible set of the CPDA*(p) problem. Then* $F \neq \emptyset$*.*

Proof. Without loss of generality, assume that p is such that $\eta_d(\mathcal{H}) = \{1,\, \ldots,\, |\eta_d(\mathcal{H})|\}$ for all $d \in \delta(\mathcal{H})$, and let $\widetilde{\mathcal{H}} := \{(d, h) \in \mathcal{H}\colon\ h \neq 1\}$. Define

$$G := \{(T,\, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}\colon\ (3.18) \text{ holds}\}, \tag{3.17}$$

where $\boldsymbol{x} = (x_{d,h},\, (d, h) \in \mathcal{H})$ and

$$x_{d,1} = \frac{[A_{d,1}(p)]^2}{T - \sum_{h \in \eta_d(\widetilde{\mathcal{H}})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} + c_d(p)}, \qquad d \in \delta(\mathcal{H}), \tag{3.18a}$$

$$\sum_{(d,h) \in \mathcal{H}} x_{d,h} = n, \tag{3.18b}$$

$$T \geq \max_{d \in \delta(\mathcal{H})} \Big(\sum_{h \in \eta_d(\widetilde{\mathcal{H}})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) + \frac{[A_{d,1}(p)]^2}{N_{d,1}} \Big), \tag{3.18c}$$

$$x_{d,h} \leq N_{d,h}, \qquad (d, h) \in \widetilde{\mathcal{H}}. \tag{3.18d}$$

Note that (3.18c) implies that the denominator in (3.18a) is strictly positive.

We prove Proposition 3.1.4 by showing that $\emptyset \neq G \subset F$.

<u>$G \subset F$.</u> Assume that $(T,\, \boldsymbol{x}) \in G$. Then $(T,\, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}$, and for this point:

- (3.2a) holds due to (3.18b);
- (3.2b) holds due to (3.18a) and the fact that $\eta_d(\mathcal{H}) = \eta_d(\widetilde{\mathcal{H}}) \cup \{1\}$ for all $d \in \delta(\mathcal{H})$;
- (3.18a) and (3.18c) jointly imply (3.2c) for $h = 1$, $d \in \delta(\mathcal{H})$; while (3.18d) is equivalent to (3.2c) for $(d, h) \in \widetilde{\mathcal{H}}$.

<u>$G \neq \emptyset$.</u> First, we consider the case $\widetilde{\mathcal{H}} \neq \emptyset$, that is, when the model p admits a domain $d \in \delta(\mathcal{H})$ with $|\eta_d(\mathcal{H})| \geq 2$. For such p, let the set K and the function k be defined as in Lemma 3.1.3.

Then, the constraints (3.18) defining the set G can equivalently be rewritten as follows:

$$x_{d,1} = \frac{[A_{d,1}(p)]^2}{T - \sum_{h \in \eta_d(\widetilde{\mathcal{H}})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} + c_d(p)}, \qquad d \in \delta(\mathcal{H}), \tag{3.19a}$$

$$k(T,\, \widetilde{\boldsymbol{x}}) = n, \tag{3.19b}$$

$$(T,\, \widetilde{\boldsymbol{x}}) \in K, \tag{3.19c}$$

where $\widetilde{\boldsymbol{x}} = (x_{d,h},\, (d, h) \in \widetilde{\mathcal{H}})$.

Because

$$k(K) \overset{(3.10)}{=} \Big(0, \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big] \overset{(2.4e)}{\ni} n, \tag{3.20}$$

we conclude that there exists $(T, \widetilde{\boldsymbol{x}}) \in K$ such that $k(T, \widetilde{\boldsymbol{x}}) = n$. Therefore, there exists $(T, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}$ satisfying (3.19). Thus, $G \neq \emptyset$.

The remaining case $\widetilde{\mathcal{H}} = \emptyset$ follows by similar arguments. Define

$$k(T) := \sum_{d\in\delta(\mathcal{H})} \frac{[A_{d,1}(p)]^2}{T + \frac{[A_{d,1}(p)]^2}{N_{d,1}}}, \qquad T \geq 0. \tag{3.21}$$

Then, for $\widetilde{\mathcal{H}} = \emptyset$, the constraints (3.18) defining the set G can equivalently be rewritten as follows:

$$x_{d,1} = \frac{[A_{d,1}(p)]^2}{T + \frac{[A_{d,1}(p)]^2}{N_{d,1}}}, \qquad d \in \delta(\mathcal{H}), \tag{3.22a}$$

$$k(T) = n, \tag{3.22b}$$

$$T \geq 0. \tag{3.22c}$$

Because

$$k([0, \infty)) = \Big(0, \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big] \overset{(2.4e)}{\ni} n, \tag{3.23}$$

we conclude that there exists $T \geq 0$ such that $k(T) = n$. Therefore, there exists $T \in \mathbb{R}$ satisfying (3.22). Thus, $G \neq \emptyset$.

□

Proposition 3.1.5. *Let $p \in \mathcal{P}$. An optimal solution to the* CPDA*(p) problem exists.*

Proof. We aim to show that the objective function $f(T, \boldsymbol{x}) := T$ attains its minimum over the feasible set $F := \{(T, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}\colon\ (3.2) \text{ holds for } p\}$, where $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n)$.

Since $F \neq \emptyset$ (Prop. 3.1.4), we may fix an arbitrary $(T_0, \boldsymbol{x}_0) \in F$ and define the sublevel set

$$F_0 := \{(T, \boldsymbol{x}) \in F\colon\ f(T, \boldsymbol{x}) \leq T_0\} = \{(T, \boldsymbol{x}) \in F\colon\ T \leq T_0\}. \tag{3.24}$$

We will show that F_0 is nonempty, bounded, and closed in the metric space $\mathbb{R}^{1+|\mathcal{H}|}$. From these properties and the continuity of f, we conclude that f attains a minimum on F.

<u>$F_0 \neq \emptyset$.</u> This is immediate, as $(T_0, \boldsymbol{x}_0) \in F_0$ by the definition of F_0.

F_0 is bounded. For any $(T,\, \boldsymbol{x}) \in F_0$, we have $T \leq T_0$ and $x_{d,h} \leq N_{d,h}$, $(d,h) \in \mathcal{H}$, so

$$\|(T,\, \boldsymbol{x})\| = \sqrt{T^2 + \sum_{(d,h)\in\mathcal{H}} x_{d,h}^2} \leq \sqrt{T_0^2 + \sum_{(d,h)\in\mathcal{H}} N_{d,h}^2}. \tag{3.25}$$

Hence, by Theorem E.1.5, F_0 is bounded in $\mathbb{R}^{1+|\mathcal{H}|}$.

F_0 is closed. From constraint (3.2b), for every $(T,\, \boldsymbol{x}) \in F_0$ and each $d \in \delta(\mathcal{H})$, we have

$$\tfrac{[A_{d,h}(p)]^2}{x_{d,h}} \leq \sum_{i\in\eta_d(\mathcal{H})} \tfrac{[A_{d,i}(p)]^2}{x_{d,i}} \leq T_0 + c_d(p), \qquad h \in \eta_d(\mathcal{H}), \tag{3.26a}$$

which implies

$$x_{d,h} \geq \tfrac{[A_{d,h}(p)]^2}{T_0 + c_d(p)} =: m_{d,h}, \qquad (d,h) \in \mathcal{H}. \tag{3.26b}$$

Since $T_0 \geq 0$ (Cor. 3.1.1) and $\frac{[A_{d,h}(p)]^2}{N_{d,h}} \leq c_d(p)$ for all $(d,h) \in \mathcal{H}$, it follows that $m_{d,h} \in (0,\, N_{d,h}]$ for all $(d,h) \in \mathcal{H}$.

Define the rectangle

$$X := [0,\, T_0] \times \Big(\bigtimes_{(d,h)\in\mathcal{H}} [m_{d,h},\, N_{d,h}] \Big), \tag{3.27}$$

and let

$$\begin{aligned} F_a &:= \{(T,\, \boldsymbol{x}) \in X\colon\ \text{(3.2a) holds for } p\}, \\ F_b &:= \{(T,\, \boldsymbol{x}) \in X\colon\ \text{(3.2b) holds for } p\}. \end{aligned} \tag{3.28}$$

Then $F_0 = F_a \cap F_b$ (recall that $(T,\, \boldsymbol{x}) \in F_0$ implies $T \geq 0$ by Cor. 3.1.1). To show that F_0 is closed in the metric space $\mathbb{R}^{1+|\mathcal{H}|}$, it suffices to show that F_a and F_b are closed, since the intersection of two closed sets is closed (see Th. E.1.2).

Define the mappings $k_1\colon X \to \mathbb{R}$ and $k_2\colon X \to \mathbb{R}^{|\delta(\mathcal{H})|}$ by:

$$\begin{aligned} k_1(T,\, \boldsymbol{x}) &:= \sum_{(d,h)\in\mathcal{H}} x_{d,h} - n, \\ k_2(T,\, \boldsymbol{x}) &:= \Big(\sum_{h\in\eta_d(\mathcal{H})} \tfrac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) - T\,,\ d \in \delta(\mathcal{H}) \Big). \end{aligned} \tag{3.29}$$

Then $F_a = k_1^{-1}(\{0\})$ and $F_b = k_2^{-1}(\{\mathbf{0}\})$.

The mappings k_1 and k_2 are continuous, and the sets $\{0\} \subset \mathbb{R}$ and $\{\mathbf{0}\} \subset \mathbb{R}^{|\delta(\mathcal{H})|}$ are closed in $\mathbb{R}$ and $\mathbb{R}^{|\delta(\mathcal{H})|}$, respectively. Thus, by Theorem E.4.2, F_a and F_b are closed in X. Since X itself is closed in $\mathbb{R}^{1+|\mathcal{H}|}$, it follows from the transitivity of closedness (Th. E.1.4) that F_a and F_b are closed in $\mathbb{R}^{1+|\mathcal{H}|}$. Consequently, their intersection $F_0 = F_a \cap F_b$ is also closed in $\mathbb{R}^{1+|\mathcal{H}|}$.

Finally, since F_0 is bounded and closed in $\mathbb{R}^{1+|\mathcal{H}|}$, the Heine-Borel Theorem (Th. E.2.1) implies that F_0 is compact in $\mathbb{R}^{1+|\mathcal{H}|}$. Given that $F_0 \neq \emptyset$ and the objective function f is continuous on F (and thus lower semicontinuous by Cor. E.4.4), the Weierstrass Theorem (Th. E.4.5, point 3) guarantees that the set of minimizers of f over F is nonempty. Consequently, an optimal solution to the CPDA(p) problem exists. □

3.2. Optimality Conditions

In this section, we derive the sufficient optimality conditions for the solution to the CPDA(p) problem, given a specific model $p \in \mathcal{P}$. These conditions, established in Theorem 3.2.8, constitute one of the central results of this dissertation. The conditions are characterized by certain set functions defined on the family of admissible sets $\mathcal{F}_p$ (recall Def. 2.2.3) and are parameterized by the model p. We begin by defining the functions and matrix structures required for this characterization.

3.2.1. The Eigen operator and the function s

Definition 3.2.1 (Matrix $\mathbf{D}_{\mathcal{A}|p}$)**.** Let $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$. Suppose that $\mathcal{A} \in \mathcal{F}_p$ is an admissible set for p, and let $m := |\delta(\mathcal{H} \setminus \mathcal{A})|$. The matrix $\mathbf{D}_{\mathcal{A}|p} \in \mathbb{R}^{m \times m}$ is defined by

$$\mathbf{D}_{\mathcal{A}|p} := \frac{1}{n - \sum_{(d,h)\in\mathcal{A}} N_{d,h}} \boldsymbol{a}\,\boldsymbol{a}^\top - \operatorname{diag}(\boldsymbol{b}), \tag{3.30}$$

where the column vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ are given by

$$\boldsymbol{a} := \Big(\sum_{h\in\eta_d(\mathcal{H}\setminus\mathcal{A})} A_{d,h}(p),\ d \in \delta(\mathcal{H} \setminus \mathcal{A}) \Big)^\top, \tag{3.31a}$$

$$\boldsymbol{b} := \Big(\sum_{h\in\eta_d(\mathcal{H}\setminus\mathcal{A})} \tfrac{[A_{d,h}(p)]^2}{N_{d,h}},\ d \in \delta(\mathcal{H} \setminus \mathcal{A}) \Big)^\top. \tag{3.31b}$$

Remark 3.2.1. Let $p \in \mathcal{P}$, $\mathcal{A} \in \mathcal{F}_p$, and denote $\delta(\mathcal{H} \setminus \mathcal{A}) = \{d_1, \ldots, d_m\}$, where $m := |\delta(\mathcal{H} \setminus \mathcal{A})|$. Then the matrix $\mathbf{D}_{\mathcal{A}|p}$ has the structure

$$\mathbf{D}_{\mathcal{A}|p} = \begin{bmatrix} \widetilde{n}\, a_{d_1}^2 - b_{d_1} & \widetilde{n}\, a_{d_1} a_{d_2} & \ldots & \widetilde{n}\, a_{d_1} a_{d_m} \\ \widetilde{n}\, a_{d_1} a_{d_2} & \widetilde{n}\, a_{d_2}^2 - b_{d_2} & \ldots & \widetilde{n}\, a_{d_2} a_{d_m} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{n}\, a_{d_1} a_{d_m} & \widetilde{n}\, a_{d_2} a_{d_m} & \ldots & \widetilde{n}\, a_{d_m}^2 - b_{d_m} \end{bmatrix}, \tag{3.32}$$

where $\widetilde{n} := \frac{1}{n - \sum_{(d,h)\in\mathcal{A}} N_{d,h}}$, and vectors $\boldsymbol{a} = (a_{d_1}, \ldots, a_{d_m})^\top$, $\boldsymbol{b} = (b_{d_1}, \ldots, b_{d_m})^\top$ are defined as in (3.31) for the given p and $\mathcal{A}$.

Following Definition 3.2.1, for a given model p, each admissible set $\mathcal{A} \in \mathcal{F}_p$ defines a matrix $\mathbf{D}_{\mathcal{A}|p}$, so that the family $\mathcal{F}_p$ induces the class of matrices

$$\big\{\mathbf{D}_{\mathcal{A}|p}\colon\ \mathcal{A} \in \mathcal{F}_p\big\}. \tag{3.33}$$

Note that while each $\mathcal{A}$ uniquely determines a matrix $\mathbf{D}_{\mathcal{A}|p}$, the mapping $\mathcal{A} \mapsto \mathbf{D}_{\mathcal{A}|p}$ is not necessarily injective; in other words, distinct admissible sets may yield the same matrix.

Lemma 3.2.1. *Let $p \in \mathcal{P}$. For every matrix $\mathbf{D}$ in the class $\big\{\mathbf{D}_{\mathcal{A}|p}\colon\ \mathcal{A} \in \mathcal{F}_p\big\}$, there exists a unique eigenpair $(\lambda,\, \boldsymbol{v}) \in \mathbb{R} \times \mathbb{R}^m$ such that*

$$\mathbf{D}\,\boldsymbol{v} = \lambda\,\boldsymbol{v}, \qquad \|\boldsymbol{v}\| = 1, \qquad \boldsymbol{v} > \mathbf{0}, \tag{3.34}$$

where m is the order of the square matrix $\mathbf{D}$. Moreover, λ is the largest eigenvalue of $\mathbf{D}$.

Proof. The following argument is based on the logic presented in Wesołowski and Wieczorkowski [45, Proof of Th. 2.1, p. 2215] and Wesołowski [44, Sec. 3, p. 9].

Let $\mathbf{D} \in \big\{\mathbf{D}_{\mathcal{A}|p}\colon\ \mathcal{A} \in \mathcal{F}_p\big\}$. Then $\mathbf{D}$ has the form (3.30) for some admissible set $\mathcal{A} \in \mathcal{F}_p$, with the vector $\boldsymbol{b} = (b_d,\, d \in \mathcal{O})^\top$ defined as in (3.31b), where $\mathcal{O} := \delta(\mathcal{H} \setminus \mathcal{A})$. Note that $\mathbf{D}$ is of order $|\mathcal{O}|$.

Let $(\lambda_d,\, \boldsymbol{v}_d)$, $d \in \mathcal{O}$, denote the eigenpairs of $\mathbf{D}$. Fix a constant $k > \max_{d\in\mathcal{O}} b_d$. Then

$$(\mathbf{D} + k\,\mathbf{I})\,\boldsymbol{v}_d = (\lambda_d + k)\,\boldsymbol{v}_d, \qquad d \in \mathcal{O}, \tag{3.35}$$

where $\mathbf{I}$ is the identity matrix of order $|\mathcal{O}|$. Hence, the eigenpairs of the shifted matrix $\mathbf{D} + k\,\mathbf{I}$ are $(\lambda_d + k,\, \boldsymbol{v}_d)$, $d \in \mathcal{O}$.

By construction, $\mathbf{D} + k\,\mathbf{I}$ is a matrix of order $|\mathcal{O}|$ with all entries strictly positive (see (3.32)). Thus, by the Perron-Frobenius Theorem (see Th. C.1), there exists exactly one $d_0 \in \mathcal{O}$ such that $(\lambda_{d_0} + k,\, \boldsymbol{v}_{d_0}) \in \mathbb{R}_+ \times \mathbb{R}^{|\mathcal{O}|}$ and:

$$\lambda_{d_0} + k \geq \max_{d\in\mathcal{O}} \big|\lambda_d + k\big|, \tag{3.36a}$$

$$\boldsymbol{v}_{d_0} > \mathbf{0}, \tag{3.36b}$$

$$\|\boldsymbol{v}_{d_0}\| = 1, \tag{3.36c}$$

$$\boldsymbol{v}_{d_0} \neq \boldsymbol{v}_d, \qquad d \in \mathcal{O} \setminus \{d_0\}. \tag{3.36d}$$

By setting $(\lambda,\, \boldsymbol{v}) := (\lambda_{d_0},\, \boldsymbol{v}_{d_0})$, we obtain a unique real eigenpair of $\mathbf{D}$ satisfying $\|\boldsymbol{v}\| = 1$ and $\boldsymbol{v} > \mathbf{0}$.

Finally, (3.36a), in view of the fact that $\lambda_d \in \mathbb{R}$ for all $d \in \mathcal{O}$ (since $\mathbf{D}$ is real and symmetric), implies that

$$\lambda_{d_0} + k \geq \max_{d \in \mathcal{O}} (\lambda_d + k), \tag{3.37}$$

which simplifies to $\lambda_{d_0} \geq \max_{d \in \mathcal{O}} \lambda_d$, confirming that λ is the largest eigenvalue of $\mathbf{D}$. □

Remark 3.2.2. For a given matrix $\mathbf{D} \in \{\mathbf{D}_{\mathcal{A}|p} \colon \ \mathcal{A} \in \mathcal{F}_p\}$, $p \in \mathcal{P}$, the eigenpair $(\lambda, \boldsymbol{v})$ satisfying the conditions (3.34) is used in the `RDCA` algorithm (presented later in Chapter 4) that solves the CPDA(p) problem. The fact that λ is the largest eigenvalue is particularly useful for the computational implementation of the `RDCA`, as it eliminates the need to verify eigenpairs against these conditions. In practice, numerical linear algebra routines (e.g., the `R` base function `eigen()`) often return eigenpairs sorted by eigenvalue magnitude, allowing the dominant eigenpair to be extracted directly without computing or inspecting all eigenpairs individually.

Definition 3.2.2 (Eigen-operator)**.** Let $p \in \mathcal{P}$. The *eigen-operator* is defined by

$$\text{Eigen}(\mathcal{A} \mid p) := (\lambda, \boldsymbol{v}), \qquad \mathcal{A} \in \mathcal{F}_p, \tag{3.38}$$

where $(\lambda, \boldsymbol{v}) \in \mathbb{R} \times \mathbb{R}^m$ is the eigenpair of the matrix $\mathbf{D}_{\mathcal{A}|p}$ (as in Def. 3.2.1) satisfying

$$\mathbf{D}_{\mathcal{A}|p}\, \boldsymbol{v} = \lambda\, \boldsymbol{v}, \qquad \|\boldsymbol{v}\| = 1, \qquad \boldsymbol{v} > \mathbf{0}, \tag{3.39}$$

and $\boldsymbol{v} = (v_d,\, d \in \delta(\mathcal{H} \setminus \mathcal{A}))$, with $m := |\delta(\mathcal{H} \setminus \mathcal{A})|$.

Remark 3.2.3. The operator Eigen defined in Definition 3.2.2 is well-defined.

Proof of Remark 3.2.3. Let $p \in \mathcal{P}$. By Lemma 3.2.1, for any $\mathcal{A} \in \mathcal{F}_p$, the associated matrix $\mathbf{D}_{\mathcal{A}|p}$ is of order $m := |\delta(\mathcal{H} \setminus \mathcal{A})|$ and possesses a unique eigenpair $(\lambda, \boldsymbol{v}) \in \mathbb{R} \times \mathbb{R}^m$ satisfying the conditions in (3.39). Consequently, Eigen$(\cdot \mid p)$ is well-defined. □

To illustrate the functioning of the Eigen operator, consider a model $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$ with $\delta(\mathcal{H}) = \{1, 2, 3\}$. Let $\mathcal{A} \in \mathcal{F}_p$ be such that $\delta(\mathcal{H} \setminus \mathcal{A}) = \{2, 3\}$. Suppose that $\big(\lambda, (w_1, w_2)\big) \in \mathbb{R} \times \mathbb{R}^2_+$ is the eigenpair of the 2×2 matrix $\mathbf{D}_{\mathcal{A}|p}$ satisfying $\|(w_1, w_2)\| = 1$. Then, Eigen$(\mathcal{A} \mid p) = (\lambda, \boldsymbol{v})$, where the vector $\boldsymbol{v} = (v_2, v_3)$ is defined by the mapping $v_2 = w_1$ and $v_3 = w_2$.

This example demonstrates that the original indexing from $\delta(\mathcal{H})$ is preserved, even when the dimension of the eigenvector $\boldsymbol{v}$ is strictly smaller than $|\delta(\mathcal{H})|$. In such cases, although $\boldsymbol{v}$ has fewer components, its entries remain indexed by the corresponding subset of $\delta(\mathcal{H})$, thereby retaining their identities relative to the full index set.

Definition 3.2.3 (Function s)**.** Let $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$. Function s is defined by

$$s(\mathcal{A},\, \boldsymbol{v},\, d \mid p) := \frac{n - \sum_{(i,h)\in\mathcal{A}} N_{i,h}}{\sum_{(i,h)\in\mathcal{H}\setminus\mathcal{A}} v_i A_{i,h}(p)}\, v_d, \tag{3.40}$$

for:

- $\mathcal{A} \in \mathcal{F}_p$,
- $\boldsymbol{v} = \big(v_i,\, i \in \delta(\mathcal{H} \setminus \mathcal{A})\big) \in \mathbb{R}_+^{|\delta(\mathcal{H}\setminus\mathcal{A})|}$,
- $d \in \delta(\mathcal{H} \setminus \mathcal{A})$.

Remark 3.2.4. The function s defined in Definition 3.2.3 is well-defined and $s > 0$.

Proof of Remark 3.2.4. Let $\mathcal{A} \in \mathcal{F}_p$. Then $n > \sum_{(d,h)\in\mathcal{A}} N_{d,h}$ (by Def. 2.2.3) and $\mathcal{H}\setminus\mathcal{A} \neq \emptyset$ (by Rem. 2.2.1). Therefore, since $v_d > 0$ for all $d \in \delta(\mathcal{H}\setminus\mathcal{A})$, the function s is well-defined and strictly positive. □

3.2.2. A relaxed auxiliary problem and its equivalence

Before formulating the optimality conditions for the CPDA problem, we introduce its relaxed counterpart, REL-CPDA, defined in Problem 3.2.1. We show that REL-CPDA is solution-equivalent to the original CPDA problem, meaning that their sets of optimal solutions coincide. The specific structure of REL-CPDA facilitates a more direct analysis of key properties which, by virtue of this equivalence, can then be extended to the original CPDA problem.

Problem 3.2.1. (REL-CPDA) For a given model $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$, the REL-CPDA($p$) optimization problem is defined as:

$$\underset{(T,\boldsymbol{x})\,\in\,\mathbb{R}\times\mathbb{R}_+^{|\mathcal{H}|}}{\text{minimize}} \quad T \tag{3.41}$$

$$\text{subject to} \quad \sum_{(d,h)\in\mathcal{H}} x_{d,h} - n = 0, \tag{3.42a}$$

$$\sum_{h\in\eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) - T \leq 0, \qquad d \in \delta(\mathcal{H}), \tag{3.42b}$$

$$x_{d,h} - N_{d,h} \leq 0, \qquad (d,h) \in \mathcal{H}, \tag{3.42c}$$

where $(T,\, \boldsymbol{x}) = \big(T,\, (x_{d,h},\, (d,h) \in \mathcal{H})\big)$ is the optimization variable.

Proposition 3.2.2. *Let $p \in \mathcal{P}$. The feasible set of the* REL-CPDA*(p) optimization problem is nonempty.*

Proof. Define

$$\begin{aligned} F &:= \{(T,\, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}\colon\ (3.2) \text{ holds for } p\}, \\ F_{rel} &:= \{(T,\, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}\colon\ (3.42) \text{ holds for } p\}. \end{aligned} \tag{3.43}$$

Since (3.42b) is a relaxation of (3.2b), we have $F \subseteq F_{rel}$. Moreover, by Proposition 3.1.4, $F \neq \emptyset$. Consequently, $F_{rel} \neq \emptyset$. □

Proposition 3.2.3. *Let $p \in \mathcal{P}$. An optimal solution to the* REL-CPDA*(p) problem exists.*

Proof. Proposition 3.2.2 ensures that the feasible set of REL-CPDA(p) is nonempty. Given this fact, the proof follows the argument of Proposition 3.1.5, with the only difference being the treatment of the set F_b.

Let the point $(T_0,\, \boldsymbol{x}_0)$, the set X, and the function k_2 be as in Prop. 3.1.5. Define

$$F_b := \{(T,\, \boldsymbol{x}) \in X\colon\ (3.42\text{b}) \text{ holds for } p\} = \{(T,\, \boldsymbol{x}) \in X\colon\ k_2(T,\, \boldsymbol{x}) \leq \mathbf{0}\}. \tag{3.44}$$

Since $(T,\, \boldsymbol{x}) \in X$ implies $T \leq T_0$ and $x_{d,h} \leq N_{d,h}$, $(d,h) \in \mathcal{H}$, it follows that

$$\big(k_2(T,\, \boldsymbol{x})\big)_d \geq \sum_{h \in \eta_d(\mathcal{H})} \left(\tfrac{[A_{d,h}(p)]^2}{x_{d,h}} - \tfrac{[A_{d,h}(p)]^2}{N_{d,h}} \right) - T_0 \geq -T_0, \qquad d \in \delta(\mathcal{H}), \tag{3.45}$$

where $\big(k_2(T,\, \boldsymbol{x})\big)_d$ denotes the dth component of the vector-valued function k_2.

Therefore,

$$F_b = k_2^{-1}([-T_0,\, 0]^{|\delta(\mathcal{H})|}). \tag{3.46}$$

Since k_2 is continuous and $[-T_0,\, 0]^{|\delta(\mathcal{H})|}$ is closed in $\mathbb{R}^{|\delta(\mathcal{H})|}$, it follows that F_b is closed in X, and consequently in $\mathbb{R}^{1+|\mathcal{H}|}$.

The remainder of the argument is identical to that of Proposition 3.1.5 and yields the existence of an optimal solution to REL-CPDA(p). □

Proposition 3.2.4. *Let $p \in \mathcal{P}$. For the* REL-CPDA*(p) problem, all inequality constraints in* (3.42b) *are active at the optimal solution $(T^*,\, \boldsymbol{x}^*)$, that is,*

$$\sum_{h \in \eta_d(\mathcal{H})} \tfrac{[A_{d,h}(p)]^2}{x^*_{d,h}} - c_d(p) - T^* = 0, \qquad d \in \delta(\mathcal{H}). \tag{3.47}$$

Proof. Let $(T^*,\, \boldsymbol{x}^*)$ be an optimal solution to the REL-CPDA(p) problem, whose existence is guaranteed by Proposition 3.2.3. Define the index sets

$$\mathcal{D}_1 := \Big\{ d \in \delta(\mathcal{H})\colon \sum_{h \in \eta_d(\mathcal{H})} \tfrac{[A_{d,h}(p)]^2}{x^*_{d,h}} - c_d(p) < T^* \Big\}, \tag{3.48a}$$

$$\mathcal{D}_2 := \delta(\mathcal{H}) \setminus \mathcal{D}_1 \ = \Big\{ d \in \delta(\mathcal{H})\colon \sum_{h \in \eta_d(\mathcal{H})} \tfrac{[A_{d,h}(p)]^2}{x^*_{d,h}} - c_d(p) = T^* \Big\}. \tag{3.48b}$$

To prove the proposition, we show that $\mathcal{D}_1 = \emptyset$.

Constraints (3.42b) and (3.42c) imply that $T^* \geq 0$. We distinguish two cases for T^*.

$T^* = 0$. Constraint (3.42c) implies that for any feasible $(T, \boldsymbol{x})$,

$$\sum_{h \in \eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) \geq 0, \qquad d \in \delta(\mathcal{H}). \tag{3.49}$$

Hence, $\mathcal{D}_1 = \emptyset$ when $T^* = 0$.

$T^* > 0$. Observe that, since the objective is to minimize T, and the constraints in (3.42b) impose lower bounds on T, at least one of these constraints must be active at optimality. Hence, $\mathcal{D}_2 \neq \emptyset$.

Assume, for contradiction, that $\mathcal{D}_1 \neq \emptyset$. We show that, in this case, one can construct a feasible point $(\widetilde{T}, \widetilde{\boldsymbol{x}})$ with $\widetilde{T} < T^*$, thereby contradicting optimality.

Choose

$$i \in \arg\max_{d \in \mathcal{D}_1} \Big(\sum_{h \in \eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x^*_{d,h}} - c_d(p) \Big) \qquad \text{and} \qquad j \in \eta_i(\mathcal{H}). \tag{3.50}$$

Because the inequality defining $\mathcal{D}_1$ is strict and $\mathcal{D}_2 \neq \emptyset$, there exist sufficiently small values

$$z \in (0,\, x^*_{i,j}), \tag{3.51a}$$

$$z_{d,h_d} \in (0,\, N_{d,h_d} - x^*_{d,h_d}], \qquad d \in \mathcal{D}_2, \tag{3.51b}$$

such that:

$$\frac{[A_{i,j}(p)]^2}{x^*_{i,j} - z} + \sum_{h \in \eta_i(\mathcal{H}) \setminus \{j\}} \frac{[A_{i,h}(p)]^2}{x^*_{i,h}} - c_i(p) < T^*, \tag{3.51c}$$

$$z = \sum_{d \in \mathcal{D}_2} z_{d,h_d}. \tag{3.51d}$$

Here, for each $d \in \mathcal{D}_2$, the index $h_d \in \{h \in \eta_d(\mathcal{H}) : x^*_{d,h} < N_{d,h}\}$ is arbitrary but fixed. The sets $\{h \in \eta_d(\mathcal{H}) : x^*_{d,h} < N_{d,h}\}$, $d \in \mathcal{D}_2$, are nonempty because, by the definition of $\mathcal{D}_2$ in the case $T^* > 0$, we have

$$0 < \sum_{h \in \eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x^*_{d,h}} - c_d(p) = \sum_{h \in \eta_d(\mathcal{H})} \Big(\frac{[A_{d,h}(p)]^2}{x^*_{d,h}} - \frac{[A_{d,h}(p)]^2}{N_{d,h}} \Big), \qquad d \in \mathcal{D}_2. \tag{3.52}$$

Given (3.51b), by the definition of $\mathcal{D}_2$, it follows that

$$\frac{[A_{d,h_d}(p)]^2}{x^*_{d,h_d} + z_{d,h_d}} + \sum_{h \in \eta_d(\mathcal{H}) \setminus \{h_d\}} \frac{[A_{d,h}(p)]^2}{x^*_{d,h}} - c_d(p) < T^*, \qquad d \in \mathcal{D}_2. \tag{3.53}$$

Independently of (3.51), by the definition of $\mathcal{D}_1$ we have

$$\sum_{h \in \eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x^*_{d,h}} - c_d(p) < T^*, \qquad d \in \mathcal{D}_1 \setminus \{i\}. \tag{3.54}$$

Now, define $(\widetilde{T}, \widetilde{\boldsymbol{x}})$ as follows. Let $\widetilde{T} \in (0, T^*)$ be such that the strict inequalities in (3.51c), (3.53), and (3.54) remain satisfied when T^* is replaced by $\widetilde{T}$; such $\widetilde{T}$ clearly exists. Let $\widetilde{\boldsymbol{x}} = (\widetilde{x}_{d,h}, (d,h) \in \mathcal{H})$ have the components

$$\widetilde{x}_{d,h} := \begin{cases} x^*_{i,j} - z, & \text{if } (d,h) = (i,j) \\ x^*_{d,h_d} + z_{d,h_d}, & \text{if } d \in \mathcal{D}_2,\ h = h_d \\ x^*_{d,h}, & \text{otherwise.} \end{cases} \tag{3.55}$$

The point $(\widetilde{T}, \widetilde{\boldsymbol{x}})$ is feasible for the REL-CPDA(p) problem as:

- $\widetilde{\boldsymbol{x}} > \mathbf{0}$ holds by construction, since $z < x^*_{i,j}$ and $z_{d,h_d} > 0$ (see (3.51a)–(3.51b));
- (3.42a) is satisfied due to $z = \sum_{d \in \mathcal{D}_2} z_{d,h_d}$ (see (3.51d));
- (3.42b) is satisfied by the choice of $\widetilde{T}$, since inequalities in (3.51c), (3.53), and (3.54) hold for the sets $\{i\}$, $\mathcal{D}_2$, and $\mathcal{D}_1 \setminus \{i\}$, respectively, whose union is $\delta(\mathcal{H})$;
- (3.42c) is satisfied by the choice of z_{d,h_d} (see (3.51b)).

Since $\widetilde{T} < T^*$, this contradicts the optimality of $(T^*, \boldsymbol{x}^*)$. Therefore, the assumption $\mathcal{D}_1 \neq \emptyset$ must be false, and we conclude that $\mathcal{D}_1 = \emptyset$.

□

Lemma 3.2.5. *For a given $p \in \mathcal{P}$, the* CPDA*(p) and* REL-CPDA*(p) problems are solution-equivalent; that is, they share the same set of optimal solutions.*

Proof. By Proposition 3.2.4, all inequality constraints (3.42b) of the REL-CPDA(p) problem are active at any optimal solution $(T^*, \boldsymbol{x}^*)$, that is, they hold as equalities. From this, we conclude that any optimal solution of REL-CPDA(p) is feasible and optimal for CPDA(p). Conversely, any optimal solution of CPDA(p) satisfies the inequalities of REL-CPDA(p) and attains the same objective value. Therefore, the two problems share exactly the same set of optimal solutions. □

Proposition 3.2.6. *Let $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$. The Karush-Kuhn-Tucker conditions* (F.11) *are necessary and sufficient for a point $(T^*, \boldsymbol{x}^*) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}$ to be a solution of the* REL-CPDA*(p) problem.*

Proof. The REL-CPDA(p) problem is a convex optimization problem because its objective function and inequality constraint functions are convex on $\mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}$, while its equality constraint functions are affine (see Def. F.2.1 and Rem. F.2.2). By Theorem F.2.4, it therefore suffices to verify the Slater condition for REL-CPDA(p) for the claim to hold.

Let F be the feasible set of the REL-CPDA(p) problem. By Proposition 3.2.2, $F \neq \emptyset$. Let $(T_0, \boldsymbol{x}_0) \in F$. Consider the inequality constraint functions related to (3.42b),

$$g_d(T, \boldsymbol{x}) := \sum_{h \in \eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) - T, \qquad (T, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}, \quad d \in \delta(\mathcal{H}). \tag{3.56}$$

For any $(T, \boldsymbol{x}_0)$ where $T > T_0$, we have $g_d(T, \boldsymbol{x}_0) < g_d(T_0, \boldsymbol{x}_0) \leq 0$ for all $d \in \delta(\mathcal{H})$. Furthermore, such a point $(T, \boldsymbol{x}_0)$ lies in $\mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|} = \operatorname{relint}(\mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|})$, which is the relative interior of the problem domain. The remaining inequality constraint functions (related to (3.42c)) are affine. This satisfies the Slater condition (F.12). Consequently, the KKT conditions (F.11) are necessary and sufficient for optimality. □

3.2.3. Optimality conditions

Theorem 3.2.7 (Karush-Kuhn-Tucker conditions for the REL-CPDA problem)**.** *Let $p \in \mathcal{P}$ and $(T^*, \boldsymbol{x}^*) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}$. Then $(T^*, \boldsymbol{x}^*)$ is a solution to the* REL-CPDA*(p) problem if and only if there exist constants $\alpha^* \in \mathbb{R}$, $\mu_d^* \geq 0$, $d \in \delta(\mathcal{H})$, and $\mu_{d,h}^* \geq 0$, $(d,h) \in \mathcal{H}$, such that*

$$1 - \sum_{d \in \delta(\mathcal{H})} \mu_d^* = 0, \tag{3.57a}$$

$$\alpha^* - \mu_d^* \frac{[A_{d,h}(p)]^2}{(x_{d,h}^*)^2} + \mu_{d,h}^* = 0, \qquad (d,h) \in \mathcal{H}, \tag{3.57b}$$

$$\sum_{(d,h) \in \mathcal{H}} x_{d,h}^* - n = 0, \tag{3.57c}$$

$$\sum_{h \in \eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x_{d,h}^*} - c_d(p) - T^* = 0, \qquad d \in \delta(\mathcal{H}), \tag{3.57d}$$

$$x_{d,h}^* - N_{d,h} \leq 0, \qquad (d,h) \in \mathcal{H}, \tag{3.57e}$$

$$\mu_{d,h}^*(x_{d,h}^* - N_{d,h}) = 0, \qquad (d,h) \in \mathcal{H}. \tag{3.57f}$$

Proof. The objective and constraint functions defining the REL-CPDA(p) problem are:

$$\begin{aligned}
f(T, \boldsymbol{x}) &:= T, \\
h(T, \boldsymbol{x}) &:= \sum_{(d,h) \in \mathcal{H}} x_{d,h} - n, \\
g_d(T, \boldsymbol{x}) &:= \sum_{h \in \eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) - T, \qquad d \in \delta(\mathcal{H}), \\
g_{d,h}(T, \boldsymbol{x}) &:= x_{d,h} - N_{d,h}, \qquad (d,h) \in \mathcal{H},
\end{aligned} \tag{3.58}$$

for $(T, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}_+^{|\mathcal{H}|}$.

Their gradients are:

$$\nabla f(T,\, \boldsymbol{x}) = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \nabla h(T,\, \boldsymbol{x}) = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \nabla g_d(T,\, \boldsymbol{x}) = \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ -\frac{[A_{d,1}(p)]^2}{x_{d,1}^2} \\ \vdots \\ -\frac{[A_{d,H_d}(p)]^2}{x_{d,H_d}^2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \nabla g_{d,h}(T,\, \boldsymbol{x}) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where $H_d := |\eta_d(\mathcal{H})|$.

By Proposition 3.2.6, the Karush-Kuhn-Tucker (KKT) conditions (F.11) are necessary and sufficient for optimality. Hence, $(T^*,\, \boldsymbol{x}^*)$ is a solution if and only if there exist KKT multipliers $\alpha^* \in \mathbb{R}$, $\mu_d^* \geq 0$, $d \in \delta(\mathcal{H})$, and $\mu_{d,h}^* \geq 0$, $(d,h) \in \mathcal{H}$, such that

$$1 - \sum_{d \in \delta(\mathcal{H})} \mu_d^* = 0, \tag{3.59a}$$

$$\alpha^* - \mu_d^* \tfrac{[A_{d,h}(p)]^2}{(x_{d,h}^*)^2} + \mu_{d,h}^* = 0, \qquad (d,h) \in \mathcal{H}, \tag{3.59b}$$

$$h(T^*,\, \boldsymbol{x}^*) = 0, \tag{3.59c}$$

$$g_d(T^*,\, \boldsymbol{x}^*) \leq 0, \qquad d \in \delta(\mathcal{H}), \tag{3.59d}$$

$$\mu_d^*\, g_d(T^*,\, \boldsymbol{x}^*) = 0, \qquad d \in \delta(\mathcal{H}), \tag{3.59e}$$

$$g_{d,h}(T^*,\, \boldsymbol{x}^*) \leq 0, \qquad (d,h) \in \mathcal{H}, \tag{3.59f}$$

$$\mu_{d,h}^*\, g_{d,h}(T^*,\, \boldsymbol{x}^*) = 0, \qquad (d,h) \in \mathcal{H}. \tag{3.59g}$$

To prove Theorem 3.2.7, it suffices to show that the systems (3.59) and (3.57) are equivalent in the following sense: for any vector

$$\xi = \big(\alpha^*,\, (\mu_d^*,\, d \in \delta(\mathcal{H})),\, (\mu_{d,h}^*,\, (d,h) \in \mathcal{H})\big) \in \mathbb{R} \times [0,\, \infty)^{|\delta(\mathcal{H})|+|\mathcal{H}|}, \tag{3.60a}$$

the following equivalence holds:

$$(T^*,\, \boldsymbol{x}^*) \text{ satisfies (3.59) with } \xi \quad \Longleftrightarrow \quad (T^*,\, \boldsymbol{x}^*) \text{ satisfies (3.57) with } \xi. \tag{3.60b}$$

Note that the two systems differ only in the conditions (3.59d)–(3.59e) and (3.57d), and let $\xi \in \mathbb{R} \times [0,\, \infty)^{|\delta(\mathcal{H})|+|\mathcal{H}|}$.

(⇒) Suppose $(T^*, \boldsymbol{x}^*)$ satisfies (3.59) with ξ. By Proposition 3.2.6, $(T^*, \boldsymbol{x}^*)$ is an optimal solution to the REL-CPDA(p) problem. Proposition 3.2.4 then implies that $g_d(T^*, \boldsymbol{x}^*) = 0$ for all $d \in \delta(\mathcal{H})$. This proves (3.57d). Since the remaining conditions in (3.57) are identical to those in (3.59), they are also satisfied for $(T^*, \boldsymbol{x}^*)$ and ξ.

(⇐) Suppose $(T^*, \boldsymbol{x}^*)$ satisfies (3.57) with ξ. Then $g_d(T^*, \boldsymbol{x}^*) = 0$ for all $d \in \delta(\mathcal{H})$. Hence, (3.59d)–(3.59e) hold. The remaining identical conditions confirm the equivalence, completing the proof. □

Theorem 3.2.8 (Optimality conditions)**.** *Let $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$. A point $(T^*, \boldsymbol{x}^*) \in \mathbb{R}^{1+|\mathcal{H}|}$ is a solution to the* CPDA*(p) problem if it takes the form*

$$T^* = \begin{cases} 0, & \text{if } \mathcal{U}^* = \mathcal{H} \\ \lambda^*, & \text{if } \mathcal{U}^* \subsetneq \mathcal{H}, \end{cases} \qquad x^*_{d,h} = \begin{cases} N_{d,h}, & (d,h) \in \mathcal{U}^* \\ s(\mathcal{U}^*, \boldsymbol{v}^*, d \mid p)\, A_{d,h}(p), & (d,h) \in \mathcal{H} \setminus \mathcal{U}^*, \end{cases} \tag{3.61}$$

where $\boldsymbol{x}^* = (x^*_{d,h}, (d,h) \in \mathcal{H})$ and the set $\mathcal{U}^* \subseteq \mathcal{H}$ is such that one of the following two cases holds:

Case 1:

$$n > \sum_{(d,h)\in\mathcal{U}^*} N_{d,h}, \tag{3.62a}$$

$$\delta(\mathcal{H} \setminus \mathcal{U}^*) = \delta(\mathcal{H}), \tag{3.62b}$$

$$(\lambda^*, \boldsymbol{v}^*) := \text{Eigen}(\mathcal{U}^* \mid p), \tag{3.62c}$$

$$\forall (d,h) \in \mathcal{H} \quad \Big((d,h) \in \mathcal{U}^* \quad \Leftrightarrow \quad s(\mathcal{U}^*, \boldsymbol{v}^*, d \mid p) \geq \tfrac{\rho_d}{S_{d,h}} \Big). \tag{3.62d}$$

Case 2:

$$\mathcal{U}^* = \mathcal{H}, \tag{3.63a}$$

$$n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}. \tag{3.63b}$$

Remark 3.2.5. In *Case 1* of Theorem 3.2.8, condition (3.62a) is equivalent to $\mathcal{U}^* \in \mathcal{F}_p$ (recall Def. 2.2.3). Therefore, by Remark 2.2.1, in this case we always have $\mathcal{U}^* \subsetneq \mathcal{H}$.

Proof of Theorem 3.2.8. By Remark 3.2.5, in *Case 1* we have $\mathcal{U}^* \in \mathcal{F}_p$; hence, Eigen$(\mathcal{U}^* \mid p)$ is well-defined. Furthermore, $\mathcal{U}^* \in \mathcal{F}_p$ along with (3.62b) and (3.62c) jointly ensure that $s(\mathcal{U}^*, \boldsymbol{v}^*, d \mid p)$ is well-defined for all $d \in \delta(\mathcal{H})$. We also recall Remark 3.2.4, which states that

$$s > 0. \tag{3.64}$$

In view of Lemma 3.2.5 and Theorem 3.2.7, to establish Theorem 3.2.8 it suffices to verify that $(T^*, \boldsymbol{x}^*) \in \mathbb{R}^{1+|\mathcal{H}|}$, with components given by (3.61), satisfies the following two conditions:

1. $\boldsymbol{x}^* \in \mathbb{R}_+^{|\mathcal{H}|}$;
2. there exist constants $\alpha^* \in \mathbb{R}$, $\mu_d^* \geq 0$ for all $d \in \delta(\mathcal{H})$, and $\mu_{d,h}^* \geq 0$ for all $(d,h) \in \mathcal{H}$, such that conditions (3.57) hold for $(T^*, \boldsymbol{x}^*)$.

Proof of 1. This follows immediately from (3.61), in view of (3.64).

Proof of 2. We first prove conditions (3.57c)–(3.57e).

(3.57c): Suppose $\mathcal{U}^* \subsetneq \mathcal{H}$ (*Case 1*). Then,

$$\begin{aligned}
\sum_{(d,h)\in\mathcal{H}} x_{d,h}^* &\overset{(3.61)}{=} \sum_{(d,h)\in\mathcal{U}^*} N_{d,h} + \sum_{(d,h)\in\mathcal{H}\setminus\mathcal{U}^*} s(\mathcal{U}^*, \boldsymbol{v}^*, d \mid p)\, A_{d,h}(p) \\
&\overset{(3.40)}{=} \sum_{(d,h)\in\mathcal{U}^*} N_{d,h} + \sum_{(d,h)\in\mathcal{H}\setminus\mathcal{U}^*} \frac{n - \sum_{(i,j)\in\mathcal{U}^*} N_{i,j}}{\sum_{(i,j)\in\mathcal{H}\setminus\mathcal{U}^*} v_i^* A_{i,j}(p)}\, v_d^*\, A_{d,h}(p) \\
&= \sum_{(d,h)\in\mathcal{U}^*} N_{d,h} + \Big(n - \sum_{(i,j)\in\mathcal{U}^*} N_{i,j}\Big) \frac{\sum_{(d,h)\in\mathcal{H}\setminus\mathcal{U}^*} v_d^*\, A_{d,h}(p)}{\sum_{(i,j)\in\mathcal{H}\setminus\mathcal{U}^*} v_i^* A_{i,j}(p)} \\
&= n.
\end{aligned} \tag{3.65}$$

For $\mathcal{U}^* = \mathcal{H}$ (*Case 2*), we trivially have:

$$\sum_{(d,h)\in\mathcal{H}} x_{d,h}^* \overset{(3.61)}{=} \sum_{(d,h)\in\mathcal{U}^*} N_{d,h} \overset{(3.63b)}{=} n. \tag{3.66}$$

(3.57d): In *Case 2*, $(T^*, \boldsymbol{x}^*) = (0, \boldsymbol{N})$ by (3.61), so condition (3.57d) holds trivially. Consider *Case 1*, where $\mathcal{H} \setminus \mathcal{U}^* \neq \emptyset$ (Rem. 3.2.5). For each $d \in \delta(\mathcal{H} \setminus \mathcal{U}^*)$, define

$$a_d := \sum_{h\in\eta_d(\mathcal{H}\setminus\mathcal{U}^*)} A_{d,h}(p) \qquad \text{and} \qquad b_d := \sum_{h\in\eta_d(\mathcal{H}\setminus\mathcal{U}^*)} \tfrac{[A_{d,h}(p)]^2}{N_{d,h}}. \tag{3.67}$$

Expanding the left-hand side of equation (3.57d) for $d \in \delta(\mathcal{H} \setminus \mathcal{U}^*)$ gives

$$\begin{aligned}
&\sum_{h\in\eta_d(\mathcal{H})} \tfrac{[A_{d,h}(p)]^2}{x_{d,h}^*} - c_d(p) - T^* \\
&\overset{(3.61)}{=} \sum_{h\in\eta_d(\mathcal{U}^*)} \tfrac{[A_{d,h}(p)]^2}{N_{d,h}} + \sum_{h\in\eta_d(\mathcal{H}\setminus\mathcal{U}^*)} \tfrac{[A_{d,h}(p)]^2}{s(\mathcal{U}^*, \boldsymbol{v}^*, d \mid p)\, A_{d,h}(p)} - \sum_{h\in\eta_d(\mathcal{H})} \tfrac{[A_{d,h}(p)]^2}{N_{d,h}} - \lambda^* \\
&\overset{(3.40)}{=} \frac{\sum_{(i,j)\in\mathcal{H}\setminus\mathcal{U}^*} v_i^* A_{i,j}(p)}{n - \sum_{(i,j)\in\mathcal{U}^*} N_{i,j}} \frac{1}{v_d^*} \sum_{h\in\eta_d(\mathcal{H}\setminus\mathcal{U}^*)} A_{d,h}(p) - \sum_{h\in\eta_d(\mathcal{H}\setminus\mathcal{U}^*)} \tfrac{[A_{d,h}(p)]^2}{N_{d,h}} - \lambda^* \\
&\overset{\substack{(2.26)\\(3.67)}}{=} \frac{\sum_{i\in\delta(\mathcal{H}\setminus\mathcal{U}^*)} v_i^* a_i}{n - \sum_{(i,j)\in\mathcal{U}^*} N_{i,j}} \frac{1}{v_d^*}\, a_d - b_d - \lambda^*.
\end{aligned} \tag{3.68}$$

Substituting (3.68) into equations in (3.57d) yields

$$\frac{1}{n-\sum_{(i,j)\in\mathcal{U}^*} N_{i,j}}\Big(\sum_{i\in\delta(\mathcal{H}\setminus\mathcal{U}^*)} v_i^* a_i\Big)\, a_d - b_d\, v_d^* = \lambda^*\, v_d^*, \qquad d\in\delta(\mathcal{H}\setminus\mathcal{U}^*). \tag{3.69}$$

Since $(\lambda^*,\, \boldsymbol{v}^*) \overset{(3.62c)}{=} \mathrm{Eigen}(\mathcal{U}^* \mid p)$, Definition 3.2.2 of the Eigen operator implies that $\boldsymbol{v}^* = \big(v_d^*,\, d\in\delta(\mathcal{H}\setminus\mathcal{U}^*)\big)$. Thus, the system (3.69) can be written as

$$\mathbf{D}\,\boldsymbol{v}^* = \lambda^*\,\boldsymbol{v}^*, \tag{3.70}$$

where

$$\begin{aligned}\mathbf{D} &:= \frac{1}{n-\sum_{(d,h)\in\mathcal{U}^*} N_{d,h}}\,\boldsymbol{a}\,\boldsymbol{a}^\top - \mathrm{diag}(\boldsymbol{b}),\\ \boldsymbol{a} &:= \big(a_d,\, d\in\delta(\mathcal{H}\setminus\mathcal{U}^*)\big)^\top,\\ \boldsymbol{b} &:= \big(b_d,\, d\in\delta(\mathcal{H}\setminus\mathcal{U}^*)\big)^\top.\end{aligned} \tag{3.71}$$

Because $\mathcal{U}^*\in\mathcal{F}_p$ in *Case 1*, the matrix $\mathbf{D}$ coincides with $\mathbf{D}_{\mathcal{U}^*|p}$ (see Def. 3.2.1). Using Definition 3.2.2 once more, we conclude that (3.70) holds, thereby establishing condition (3.57d) for all $d\in\delta(\mathcal{H}\setminus\mathcal{U}^*) \overset{(3.62b)}{=} \delta(\mathcal{H})$.

<u>(3.57e):</u> Note that $\frac{N_{d,h}}{A_{d,h}(p)} = \frac{\rho_d}{S_{d,h}}$ for all $(d,h)\in\mathcal{H}$. Then,

$$x_{d,h}^* \overset{(3.61)}{=} \begin{cases} N_{d,h}, & (d,h)\in\mathcal{U}^*\\ s(\mathcal{U}^*,\, \boldsymbol{v}^*,\, d\mid p)\, A_{d,h}(p) \overset{(3.62d)}{<} N_{d,h}, & (d,h)\in\mathcal{H}\setminus\mathcal{U}^*.\end{cases} \tag{3.72}$$

To prove the remaining conditions in (3.57), recall (3.64) and define the multipliers as follows:

$$\begin{aligned}\alpha^* &:= \begin{cases}\frac{1}{\sum_{d\in\delta(\mathcal{H})} s^2(\mathcal{U}^*,\boldsymbol{v}^*,d\,|\,p)}, & \text{if } \mathcal{U}^*\subsetneq\mathcal{H}\\ k, & \text{if } \mathcal{U}^*=\mathcal{H},\end{cases}\\ \mu_d^* &:= \begin{cases} s^2(\mathcal{U}^*,\, \boldsymbol{v}^*,\, d\mid p)\,\alpha^*, & \text{if } \mathcal{U}^*\subsetneq\mathcal{H}\\ k_d, & \text{if } \mathcal{U}^*=\mathcal{H},\end{cases} \qquad d\in\delta(\mathcal{H}),\\ \mu_{d,h}^* &:= \begin{cases}\mu_d^*\,\frac{S_{d,h}^2}{\rho_d^2} - \alpha^*, & (d,h)\in\mathcal{U}^*\\ 0, & (d,h)\in\mathcal{H}\setminus\mathcal{U}^*,\end{cases}\end{aligned} \tag{3.73}$$

where $k\in\mathbb{R}$ and $k_d\geq 0$, $d\in\delta(\mathcal{H})$, are arbitrary constants satisfying

$$\sum_{i\in\delta(\mathcal{H})} k_i = 1 \qquad \text{and} \qquad k_d\frac{S_{d,h}^2}{\rho_d^2} - k \geq 0,\ (d,h)\in\mathcal{H}. \tag{3.74}$$

The existence of such constants is evident. We first verify that all multipliers μ_d^* and $\mu_{d,h}^*$ are nonnegative.

$\underline{\mu_d^* \geq 0.}$ This follows directly from (3.73).

$\underline{\mu_{d,h}^* \geq 0.}$ For $(d,h) \in \mathcal{U}^*$, we have:

$$\begin{aligned}\mu_{d,h}^* &= \begin{cases} s^2(\mathcal{U}^*,\, \boldsymbol{v}^*,\, d \mid p)\, \alpha^* \frac{S_{d,h}^2}{\rho_d^2} - \alpha^*, & \text{if } \mathcal{U}^* \subsetneq \mathcal{H} \\ k_d \frac{S_{d,h}^2}{\rho_d^2} - \alpha^*, & \text{if } \mathcal{U}^* = \mathcal{H}, \end{cases} \\ &= \begin{cases} \alpha^* \Big(s^2(\mathcal{U}^*,\, \boldsymbol{v}^*,\, d \mid p)\, \frac{S_{d,h}^2}{\rho_d^2} - 1 \Big), & \text{if } \mathcal{U}^* \subsetneq \mathcal{H} \\ k_d \frac{S_{d,h}^2}{\rho_d^2} - k, & \text{if } \mathcal{U}^* = \mathcal{H}. \end{cases}\end{aligned} \tag{3.75}$$

In the case $\mathcal{U}^* \subsetneq \mathcal{H}$, non-negativity follows from condition (3.62d), which ensures

$$s^2(\mathcal{U}^*,\, \boldsymbol{v}^*,\, d \mid p)\, \frac{S_{d,h}^2}{\rho_d^2} \geq 1. \tag{3.76}$$

The remaining case is ensured by the selection of k_d and k, as specified in (3.74).

Finally, we verify that the remaining conditions in (3.57) are satisfied for the chosen multipliers.

<u>(3.57a):</u>

$$\begin{aligned}&\sum_{d \in \delta(\mathcal{H})} \mu_d^* \\ &= \begin{cases} \sum_{d \in \delta(\mathcal{H})} s^2(\mathcal{U}^*,\, \boldsymbol{v}^*,\, d \mid p)\, \frac{1}{\sum_{i \in \delta(\mathcal{H})} s^2(\mathcal{U}^*,\, \boldsymbol{v}^*,\, i \mid p)} = 1, & \text{if } \mathcal{U}^* \subsetneq \mathcal{H} \\ \sum_{d \in \delta(\mathcal{H})} k_d = 1, & \text{if } \mathcal{U}^* = \mathcal{H}. \end{cases}\end{aligned} \tag{3.77}$$

<u>(3.57b):</u> In view of (3.61),

$$\begin{aligned}&\alpha^* - \mu_d^* \tfrac{[A_{d,h}(p)]^2}{(x_{d,h}^*)^2} + \mu_{d,h}^* \\ &= \begin{cases} \alpha^* - \mu_d^* \frac{[A_{d,h}(p)]^2}{N_{d,h}^2} + \mu_d^* \frac{S_{d,h}^2}{\rho_d^2} - \alpha^* = 0, & (d,h) \in \mathcal{U}^* \\ \alpha^* - s^2(\mathcal{U}^*,\, \boldsymbol{v}^*,\, d \mid p)\, \alpha^* \frac{[A_{d,h}(p)]^2}{\Big[s(\mathcal{U}^*,\, \boldsymbol{v}^*,\, d \mid p)\, A_{d,h}(p) \Big]^2} = 0, & (d,h) \in \mathcal{H} \setminus \mathcal{U}^*. \end{cases}\end{aligned} \tag{3.78}$$

<u>(3.57f):</u> Holds trivially because

$$\begin{aligned} x_{d,h}^* - N_{d,h} &= 0, \qquad (d,h) \in \mathcal{U}^*, \\ \mu_{d,h}^* &= 0, \qquad (d,h) \in \mathcal{H} \setminus \mathcal{U}^*. \end{aligned} \tag{3.79}$$

□

According to Theorem 3.2.8, the pair $(\lambda^*, \boldsymbol{v}^*)$ plays a crucial role in the optimality conditions for the CPDA(p) problem when $\mathcal{U}^* \subsetneq \mathcal{H}$. By the definition of the Eigen operator and due to condition (3.62b), the vector $\boldsymbol{v}^*$ lies in $\mathbb{R}_+^{|\delta(\mathcal{H})|}$. In Corollary 3.2.9 below, we show that $\lambda^* > 0$.

Corollary 3.2.9. *In* Case 1 *of Theorem 3.2.8, the pair* $(\lambda^*, \boldsymbol{v}^*)$ *satisfies* $\lambda^* > 0$.

Proof. By (3.61), we have

$$x^*_{d,h} = s(\mathcal{U}^*, \boldsymbol{v}^*, d \mid p)\, A_{d,h}(p), \qquad (d,h) \in \mathcal{H} \setminus \mathcal{U}^*. \tag{3.80}$$

Furthermore, condition (3.62d) implies

$$s(\mathcal{U}^*, \boldsymbol{v}^*, d \mid p) < \tfrac{\rho_d}{S_{d,h}} = \tfrac{N_{d,h}}{A_{d,h}(p)}, \qquad (d,h) \in \mathcal{H} \setminus \mathcal{U}^*. \tag{3.81}$$

Combining these two, we obtain

$$x^*_{d,h} < N_{d,h}, \qquad (d,h) \in \mathcal{H} \setminus \mathcal{U}^*. \tag{3.82}$$

By Remark 3.2.5, in *Case 1*, $\mathcal{H} \setminus \mathcal{U}^* \neq \emptyset$, so $\boldsymbol{x}^* \neq \boldsymbol{N}$. It then follows from Corollary 3.1.1 that $T^* > 0$, which by (3.61) is implies $\lambda^* > 0$. □

Theorem 3.2.8 provides sufficient conditions for a point $(T^*, \boldsymbol{x}^*) \in \mathbb{R}^{1+|\mathcal{H}|}$ to be a solution to the CPDA problem, expressed in terms of a set $\mathcal{U}^* \subseteq \mathcal{H}$. The `RDCA` algorithm developed in Chapter 4 always finds a solution of this form by identifying a suitable $\mathcal{U}^*$ (see Theorem 6.4.2 in Chapter 6). Because the algorithm always returns a point satisfying these sufficient optimality conditions, we can conclude that the set of optimal solutions of this form is nonempty – something that could not be inferred from the conditions in Theorem 3.2.8 alone. In this sense, the algorithm provides a constructive existence proof for such solutions.

Chapter 4

The RDCA Algorithm

In this chapter, we present the *Recursive Domain-Controlled Allocation* (`RDCA`) algorithm, designed to solve the CPDA problem. The algorithm is based on the sufficient optimality conditions of Theorem 3.2.8, which characterize an optimal solution in terms of a particular subset $\mathcal{U}^* \subseteq \mathcal{H}$, referred to as the *optimal take-max* set. This subset is not known a priori. Accordingly, the primary task of `RDCA` is to identify the optimal set $\mathcal{U}^*$.

To this end, `RDCA` employs the *Domain-Controlled Allocation* (`DCA`) algorithm as its base procedure. For a given set $\mathcal{U} \subseteq \mathcal{H}$, `DCA` computes a candidate solution $(T, \boldsymbol{x})$ in the form specified by the optimality conditions (3.61), without verifying whether all conditions of either *Case 1* or *Case 2* are satisfied. This verification is performed by `RDCA` itself. If the candidate solution satisfies all required conditions, the algorithm terminates; otherwise, the set $\mathcal{U}$ is updated to account for any violations, and the process continues. As indicated in Chapter 1, the general approach of iteratively constructing an optimal solution until the optimality conditions are satisfied was motivated by the `RNA` algorithm, which solves the special case of the CPDA(p) problem for the model $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n)$ when $|\delta(\mathcal{H})| = 1$.

The `DCA`$(p, \mathcal{U})$ algorithm extends the existing `DCA`$(p, \emptyset)$ procedure, which is described in more detail in Appendix G. In this more general formulation, the input subset $\mathcal{U} \subseteq \mathcal{H}$ may be any set that satisfies the required conditions of the `DCA`$(p, \mathcal{U})$, thereby enabling its iterative application within the `RDCA` framework.

To illustrate `RDCA` in practice, we provide numerical examples that highlight the algorithm's recursive structure and how the optimal set $\mathcal{U}^*$ is determined (see Sec. 4.2).

The formal proof of `RDCA`'s correctness is given in Chapter 6.

4.1. Algorithm Definition

The definition of the `RDCA` algorithm is given as pseudocode in Algorithm 1. `RDCA` is defined recursively over the domain set $\mathcal{J} \subseteq \delta(\mathcal{H})$, utilizing `DCA` as its base case. `DCA` computes a tentative solution $(T, \boldsymbol{x})$ for a given candidate set $\mathcal{U} \subseteq \mathcal{H}$ according to (3.61). We note that although the selection of j on line 2 is non-deterministic, we show in Chapter 6 that this choice does not affect the optimality of the computed solution.

Algorithm 1 `RDCA`

Input: $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$, $\mathcal{U} \subseteq \mathcal{H}$, $\mathcal{J} \subseteq \delta(\mathcal{H})$. ▷ Def. 2.2.2 and 2.3.1

Require: $\big((n > \sum_{(d,h)\in\mathcal{U}} N_{d,h})$ or $(\mathcal{U} = \mathcal{H}$ and $n = \sum_{(d,h)\in\mathcal{H}} N_{d,h})\big)$ and $\delta(\mathcal{U}) \cap \mathcal{J} = \emptyset$ and $\mathcal{J} \neq \emptyset$.

1: **function** `RDCA`$(p, \mathcal{U}, \mathcal{J})$
2: $\quad j \in \mathcal{J}$ ▷ chosen arbitrarily
3: $\quad \mathcal{K} \leftarrow \{(d,h) \in \mathcal{H}\colon\ d = j\}$
4: $\quad$ **do**
5: $\quad\quad$ **if** $|\mathcal{J}| = 1$ **then**
6: $\quad\quad\quad (T, \boldsymbol{x}) \leftarrow$ `DCA`$(p, \mathcal{U})$
7: $\quad\quad$ **else**
8: $\quad\quad\quad (T, \boldsymbol{x}) \leftarrow$ `RDCA`$(p, \mathcal{U}, \mathcal{J} \setminus \{j\})$
9: $\quad\quad$ **end if**
10: $\quad\quad \mathcal{Y} \leftarrow \big\{(d,h) \in \mathcal{K}\colon\ x_{d,h} \geq N_{d,h}\big\}$ ▷ $\boldsymbol{x} = \big(x_{d,h},\ (d,h) \in \mathcal{H}\big)$
11: $\quad\quad$ **if** $\mathcal{Y} \neq \emptyset$ **then**
12: $\quad\quad\quad \mathcal{U} \leftarrow \mathcal{U} \cup \mathcal{Y}$
13: $\quad\quad\quad \mathcal{K} \leftarrow \mathcal{K} \setminus \mathcal{Y}$
14: $\quad\quad$ **end if**
15: $\quad$ **while** $\mathcal{Y} \neq \emptyset$
16: $\quad$ **return** $(T, \boldsymbol{x})$
17: **end function**

`RDCA` takes three arguments: the model $p \in \mathcal{P}$ and the auxiliary parameters $\mathcal{U} \subseteq \mathcal{H}$ and $\mathcal{J} \subseteq \delta(\mathcal{H})$. Parameters $\mathcal{U}$ and $\mathcal{J}$ facilitate the algorithm's recursive structure. For a given model $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$, the CPDA$(p)$ problem is solved by `RDCA`$(p, \mathcal{U}, \mathcal{J})$ with $\mathcal{U} = \emptyset$ and $\mathcal{J} = \delta(\mathcal{H})$. Numerical examples in Section 4.2 provide further illustration.

Algorithm 2 `DCA`

Input: $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$, $\mathcal{U} \subseteq \mathcal{H}$. ▷ Def. 2.2.2

Require: $(n > \sum_{(d,h)\in\mathcal{U}} N_{d,h})$ or $(\mathcal{U} = \mathcal{H}$ and $n = \sum_{(d,h)\in\mathcal{H}} N_{d,h})$.

1: **function** `DCA`$(p,\, \mathcal{U})$
2: **if** $\mathcal{U} = \mathcal{H}$ **then**
3: $T \leftarrow 0$
4: **else**
5: $(\lambda,\, \boldsymbol{v}) \leftarrow \mathrm{Eigen}(\mathcal{U} \mid p)$ ▷ Def. 3.2.2
6: $T \leftarrow \lambda$
7: **end if**
8: **return** $(T,\, \boldsymbol{x})$ where $x_{d,h} = \begin{cases} N_{d,h}, & (d,h) \in \mathcal{U} \\ s(\mathcal{U},\, \boldsymbol{v},\, d \mid p)\, A_{d,h}(p), & (d,h) \in \mathcal{H} \setminus \mathcal{U} \end{cases}$ ▷ Def. 3.2.3
9: **end function**

As indicated in the introduction of this chapter, the key task of `RDCA`$(p,\, \emptyset,\, \delta(\mathcal{H}))$ is to identify the unknown set $\mathcal{U}^* \subseteq \mathcal{H}$ corresponding to the optimal solution $(T^*,\, \boldsymbol{x}^*)$ of the CPDA(p) problem for a given $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$. The algorithm achieves this by iteratively and recursively refining a candidate set $\mathcal{U} \subseteq \mathcal{H}$, guided by the sufficient optimality conditions of Theorem 3.2.8. At the base case of the recursion, `DCA` is used to compute a pair $(T,\, \boldsymbol{x})$ in the form specified by (3.61) for a given set $\mathcal{U}$. This pair is then propagated to higher recursion levels, where violations of the inequality constraints $\boldsymbol{x} \leq \boldsymbol{N}$ are checked. If no violations occur at any recursion level, $(T,\, \boldsymbol{x})$ constitutes a solution; otherwise, the set $\mathcal{U}$ is updated and the procedure is repeated.

To handle multiple domains, `RDCA` employs recursion. When $|\mathcal{J}| \geq 2$, the algorithm selects a domain $j \in \mathcal{J}$ and recursively invokes the algorithm on the reduced set $\mathcal{J} \setminus \{j\}$. This produces an allocation that satisfies $\boldsymbol{x} \leq \boldsymbol{N}$ for all domains in $\mathcal{J}\setminus\{j\}$, but may violate these constraints within domain j. Such violations are resolved iteratively by including the corresponding strata in the set $\mathcal{U}$ and recomputing the allocation until all inequality constraints for domain j are satisfied. By construction, each level of recursion resolves violations in a single domain, ensuring that upon termination, no violations remain in any domain of $\mathcal{J}$.

Consider the set $\mathcal{U}$ that is passed to `DCA` when it is invoked, that is, at the recursion level where $|\mathcal{J}| = 1$. Following the definition of the `DCA` algorithm (see line 8), this set

can be viewed as the collection of strata $(d,h) \in \mathcal{H}$ that are *blocked* at $N_{d,h}$. During the execution of the algorithm, it may occur that, at this base-case level, all strata of a given domain belong to $\mathcal{U}$; this situation is referred to as a *domain blockage* (consider Rem. 2.3.1 for $\mathcal{B} = \mathcal{U}$). This case is illustrated by the numerical example in Section 4.2.1 and analyzed in detail in Section 6.2 of Chapter 6.

Observe that, at any recursion level, a set $\mathcal{U}$ containing all strata of some domain cannot coincide with the optimal set $\mathcal{U}^*$ when $n < \sum_{(d,h)\in\mathcal{H}} N_{d,h}$ and `RDCA` is initially invoked with $\mathcal{J} = \delta(\mathcal{H})$. Indeed, in this case we would have $T^* = 0$ (by constraint (3.2b)), and consequently, Corollary 3.1.1 implies $\boldsymbol{x}^* = \boldsymbol{N}$. This would yield $n \overset{(3.2a)}{=} \sum_{(d,h)\in\mathcal{H}} x^*_{d,h} = \sum_{(d,h)\in\mathcal{H}} N_{d,h}$, contradicting $n < \sum_{(d,h)\in\mathcal{H}} N_{d,h}$.

An important aspect of the `RDCA` algorithm is that, by checking for violations of the inequality $\boldsymbol{x} \leq \boldsymbol{N}$, it implicitly ensures that all conditions defining the optimal set $\mathcal{U}^*$ – namely, (3.62) or (3.63) – are satisfied. A formal proof of this implication is provided in Chapter 6.

For completeness, we indicate that the well-definedness and termination of `RDCA` are established in Proposition 5.2.2 in the subsequent Chapter 5. Specifically, we prove that every recursive invocation of `RDCA` or `DCA` satisfies the specified input requirements.

4.2. Examples

To illustrate the `RDCA` algorithm, we present three numerical examples using artificial populations with two and three domains, comprising four and six strata, respectively.

The numerical results in this section were obtained using R with the `rdca()` and `dca()` functions from the **stratallo** package, which implement the `RDCA` and `DCA` algorithms, respectively (see Sec. 4.3). For reproducibility, we note that R represents numbers in double-precision floating-point format; consequently, all computations are performed with finite precision rather than symbolic arithmetic. Floating-point values reported in the tables are rounded for presentation purposes only, with rounding applied after all computations are completed.

In this context, since $\boldsymbol{\rho}$ and its component-wise square $\boldsymbol{\rho}^2$ are arguments to the `rdca()` and `dca()` functions, we clarify that $\boldsymbol{\rho}$ is computed as $\boldsymbol{t} \circ \sqrt{\boldsymbol{\kappa}}$, and $\boldsymbol{\rho}^2$ is then obtained directly from $\boldsymbol{\rho}$ rather than as $\boldsymbol{t}^2 \circ \boldsymbol{\kappa}$ (where $\circ$ denotes the Hadamard or component-wise product).

4.2.1. Two domains

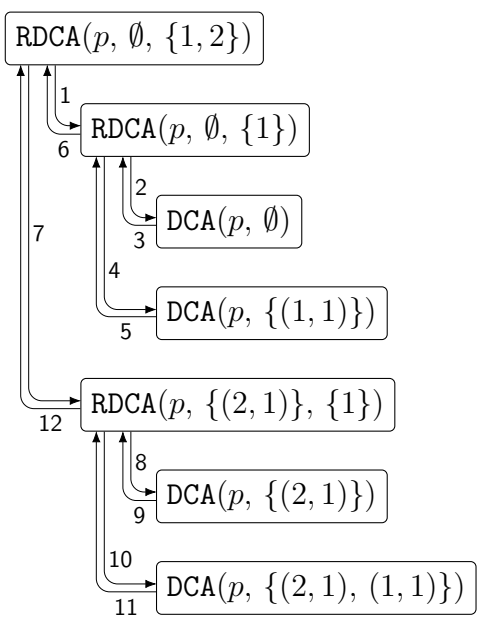
We first consider a population with two domains and four strata, and examine the allocations for two different total sample sizes, as detailed in Table 4.2.1.

Table 4.2.1: Optimum allocations for two total sample sizes across two domains and four strata.

$\delta(\mathcal{H})$	$\mathcal{H}$	$\boldsymbol{N}$	$\boldsymbol{S}$	$\boldsymbol{t}$	$\boldsymbol{\kappa}$	$\boldsymbol{\rho} = \boldsymbol{t} \circ \sqrt{\boldsymbol{\kappa}}$	$\boldsymbol{x}^*$	$\boldsymbol{x}^*$
1	(1,1)	100	10	2	0.4	1.3	100	100
	(1,2)	200	2				65.4	131.5
2	(2,1)	150	50	3	0.6	2.3	150	150
	(2,2)	40	30				34.6	38.5
total sample size n							350	420

The invocation hierarchy of the `RDCA` algorithm for the population in Table 4.2.1 and total sample size $n = 350$ is illustrated in Figure 4.2.1.

Figure 4.2.1: Recursive invocation diagram of `RDCA` for the model $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, 350)$ given in Table 4.2.1. Boxes represent calls to `RDCA` or its base case `DCA`, with arrows indicating the order of calls and returns. The table on the right reports the output $(T, \boldsymbol{x})$ for each invocation, indexed by the corresponding step number. Blue cells indicate *take-max* allocations (i.e., full allocation of the stratum size), while red cells indicate allocations exceeding the stratum population size.



step no.	T	$x_{1,1}$	$x_{1,2}$	$x_{2,1}$	$x_{2,2}$
$(T^*, \boldsymbol{x}^*)$	←12				
6	←5				
3	575.6	119.4	47.8	157.6	25.2
5	967.3	100	68.2	156.8	25.1
12	←11				
9	723.4	117.1	46.8	150	36.1
11	1030.1	100	65.4	150	34.6

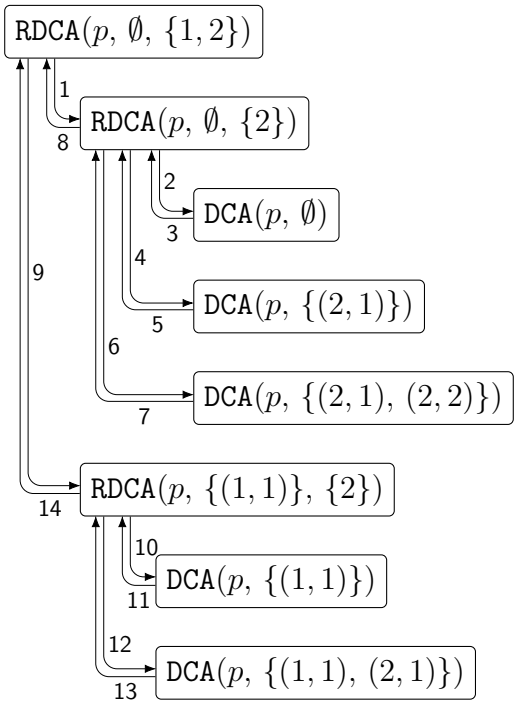
In the outermost invocation of `RDCA`, domain $j = 2$ is selected from the set $\mathcal{J} = \{1, 2\}$. The `do-while` loop iterates twice, resulting in two recursive `RDCA` calls (step numbers `1` and `7` in Fig. 4.2.1). These recursive calls ensure that the inequality constraints (3.2c) are satisfied for all strata in domain 1. As the recursion proceeds, the following four sets arise as candidates for the optimal set $\mathcal{U}^*$:

$$\emptyset, \qquad \{(1,1)\}, \qquad \{(2,1)\}, \qquad \{(2,1),\ (1,1)\}.$$

For each candidate set, a tentative allocation is computed using `DCA`, as illustrated in the diagram. Since the allocation corresponding to $\mathcal{U} = \{(2,1),\ (1,1)\}$ satisfies all inequality constraints, this set is identified as the optimal set $\mathcal{U}^*$.

As a second example, we consider the allocation for the population in Table 4.2.1 and total sample size $n = 420$. The `RDCA` invocation diagram is shown in Figure 4.2.2.

Figure 4.2.2: Recursive invocation diagram of `RDCA` for the model $p = (\mathcal{H},\ \boldsymbol{N},\ \boldsymbol{S},\ \boldsymbol{\rho},\ 420)$ given in Table 4.2.1. Boxes represent calls to `RDCA` or its base case `DCA`, with arrows indicating the order of calls and returns. The table on the right reports the output $(T,\ \boldsymbol{x})$ for each invocation, indexed by the corresponding step number. Blue cells indicate *take-max* allocations (i.e., full allocation of the stratum size), while red cells indicate allocations exceeding the stratum population size.



step no.	T	$x_{1,1}$	$x_{1,2}$	$x_{2,1}$	$x_{2,2}$
$(T,\ \boldsymbol{x}^*)$	←14				
8	←7				
3	−1474.4	165.9	66.3	161.9	25.9
5	−1210.2	157.9	63.2	150	48.9
7	−1423.9	164.3	65.7	150	40
14	←13				
11	233.1	100	136.4	158.3	25.3
13	260.4	100	131.5	150	38.5

In this example, the outermost `RDCA` invocation selects domain $j = 1$ from the set $\mathcal{J} = \{1, 2\}$. This invocation performs two iterations of the `do-while` loop, resulting in two recursive `RDCA` calls (step numbers `1` and `9` in Fig. 4.2.2). These recursive calls manage the allocation to ensure that the inequality constraints (3.2c) are satisfied for strata in domain 2, while the outermost invocation handles strata in domain 1. The sequence of candidate sets for the optimal $\mathcal{U}^*$ is:

$$\emptyset, \qquad \{(2,1)\}, \qquad \{(2,1),\, (2,2)\}, \qquad \{(1,1)\}, \qquad \{(1,1),\, (2,1)\}.$$

The allocation corresponding to $\mathcal{U} = \{(1,1),\, (2,1)\}$ satisfies all inequality constraints; hence, this set is identified as the optimal set $\mathcal{U}^*$.

Importantly, this example differs from the previous one in that a *domain-blockage* scenario occurs. Specifically, during the second iteration of the `do-while` loop in the `RDCA`$(p, \emptyset, \{2\})$ invocation, the allocation from `DCA`$(p, \{(2,1)\})$ is $\boldsymbol{x} = (157.9,\, 63.2,\, 150,\, 48.9)$, where $x_{2,2} = 48.9 > 40 = N_{2,2}$ (see step `5`). Consequently, `RDCA`$(p, \emptyset, \{2\})$ appends stratum $(2,2)$ to its set $\mathcal{U}$, resulting in all strata from domain 2 being included in $\mathcal{U}$; that is, the entire domain 2 is blocked. In the subsequent iteration, `DCA`$(p, \{(2,1),\, (2,2)\})$ computes the allocation for all domains, in particular yielding $x_{2,1} = N_{2,1}$ and $x_{2,2} = N_{2,2}$ (see step `7`). This is the final iteration of the loop, and hence, `RDCA`$(p, \emptyset, \{2\})$ returns (step `8`) the allocation in which domain 2 is allocated at its maximum size. Note that the outermost invocation, `RDCA`$(p, \emptyset, \{1,2\})$, proceeds to its second iteration of the loop because the allocation for stratum $(1,1)$, $x_{1,1} = 164.3$, exceeds the stratum's size. Finally, the optimal $\mathcal{U}^* = \{(1,1),\, (2,1)\}$; hence, no domain is blocked. The optimal allocation for domain 2 is $x^*_{2,1} = N_{2,1} = 150$ and $x^*_{2,2} = 38.5 < N_{2,2}$; that is, only part of domain 2 reaches its maximum allocation.

4.2.2. Three domains

To illustrate the `RDCA` algorithm for a population with three domains, we consider the population described in Table 4.2.2. The corresponding `RDCA` function invocation diagram is shown in Figure 4.2.3.

Table 4.2.2: Optimum allocation for a population with three domains and six strata.

$\delta(\mathcal{H})$	$\mathcal{H}$	$\boldsymbol{N}$	$\boldsymbol{S}$	$\boldsymbol{t}$	$\boldsymbol{\kappa}$	$\boldsymbol{\rho} = \boldsymbol{t} \circ \sqrt{\boldsymbol{\kappa}}$	$\boldsymbol{x}^*$
1	(1,1)	100	10	2	0.3	1.1	89.1
	(1,2)	100	30				100
2	(2,1)	100	50	3	0.2	1.3	100
	(2,2)	100	10				84.5
3	(3,1)	100	50	9	0.5	6.4	95.0
	(3,2)	100	6				11.4
total sample size n							480

As already pointed out, in the `RDCA` algorithm, each recursive invocation selects a domain j from the current set $\mathcal{J}$ of available domains. This selection is independent within the current set of available domains, although the set $\mathcal{J}$ itself is constrained by choices made in higher levels of recursion. Consequently, different branches of the recursion tree may correspond to different domain selections.

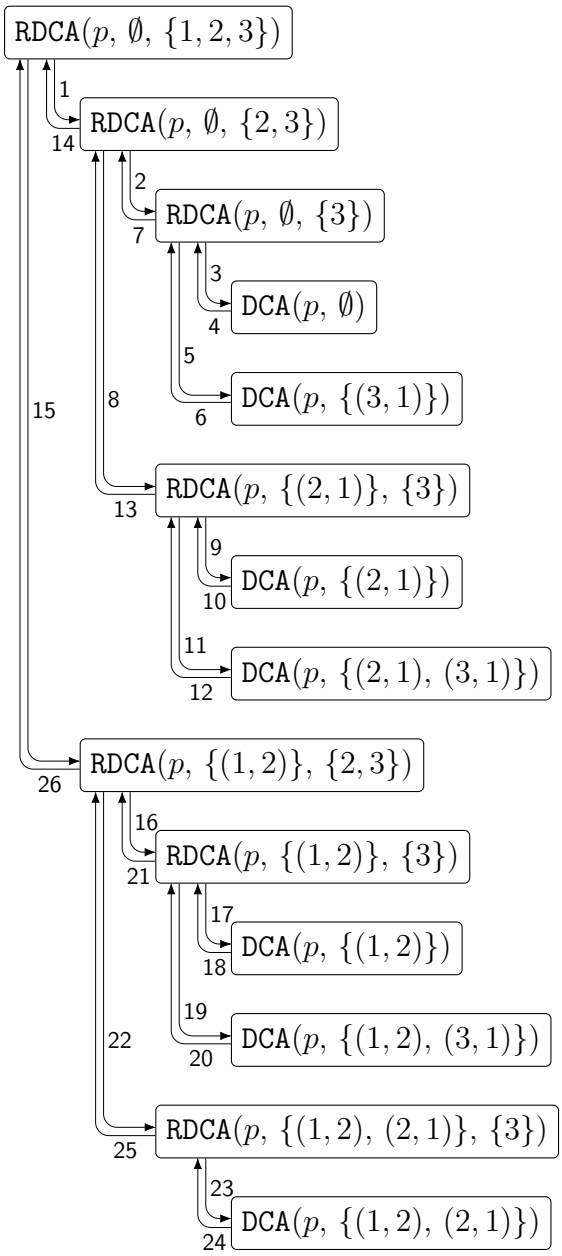
The example shown in Figure 4.2.3 illustrates one such execution path: in the outermost recursion, domain $j = 1$ is selected, followed by domain $j = 2$ at the next recursive level, and finally domain $j = 3$ in the innermost recursion. This represents one possible trajectory through the algorithm's recursive space and serves as a single example among various potential domain selection sequences. Specifically, in the invocation `RDCA`$(p, \emptyset, \{2, 3\})$, domain $j = 2$ was chosen from the set $\{2, 3\}$; however, domain $j = 3$ could have been selected instead.

4.3. Implementation in R

The `RDCA` algorithm is implemented in R as the function `rdca()` in the **stratallo** package. The `DCA` algorithm is also available in the same package as the function `dca()`. This package, developed and maintained by the author, provides functions for various optimum allocation problems (see Wójciak [51]) and is available on the Comprehensive R Archive Network (CRAN) [35]. See also Remark 3.2.2 for a relevant implementation detail concerning the `DCA` procedure.

The source code of the **stratallo** package is hosted on GitHub (see Wójciak, Wesołowski and Wieczorkowski [52]).

Figure 4.2.3: Recursive invocation diagram of `RDCA` for the model $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, 480)$ given in Table 4.2.2. Boxes represent calls to `RDCA` or its base case `DCA`, with arrows indicating the order of calls and returns. The table on the right reports the output $(T, \boldsymbol{x})$ for each invocation, indexed by the corresponding step number. Blue cells indicate *take-max* allocations (i.e., full allocation of the stratum size), while red cells indicate allocations exceeding the stratum population size.



step no.	T	$x_{1,1}$	$x_{1,2}$	$x_{2,1}$	$x_{2,2}$	$x_{3,1}$	$x_{3,2}$
$(T, \boldsymbol{x}^*)$	←26						
14	←13						
7	←6						
4	−1863.9	40.9	122.7	116.9	23.4	157.2	18.9
6	20.0	40.0	120.0	115.4	23.1	100	81.6
13	←12						
10	93.6	40.0	119.9	100	98.3	108.8	13.1
12	265.6	39.9	119.6	100	95.4	100	25.1
26	←25						
21	←20						
18	−507.5	106.5	100	115.8	23.2	120.1	14.4
20	117.6	98.6	100	115.3	23.1	100	43.0
25	←24						
24	1017.8	89.1	100	100	84.5	95.0	11.4

Chapter 5

Structural Properties of the RDCA Algorithm

This chapter and the one following are dedicated to establishing that the `RDCA` algorithm solves the CPDA problem. While the ultimate proof of total correctness is presented in Chapter 6, the current chapter provides the necessary groundwork and foundational structural characterization of `RDCA` upon which the proof of correctness is built.

Section 5.1 introduces notational conventions for referencing the values of program variables across loop iterations and recursive invocations. These conventions provide the formal foundation for the entire analytical development of both this and the following chapter.

In Section 5.2, we show that `RDCA` is well-defined and guaranteed to terminate. Well-definedness means that at each iteration of the loop, the input parameters to `DCA` and to recursive invocations of `RDCA` satisfy all required preconditions, while termination ensures that the algorithm halts after a finite number of steps.

Section 5.3 describes fundamental relationships among the variables of `RDCA`, which are used repeatedly throughout the proofs in Chapter 6.

Section 5.4 defines the *Last-Branch Recursion Path*, a sequence of recursive `RDCA` invocations that are pivotal to proving the correctness of the algorithm.

Section 5.5 defines a sequence of take-max strata sets $\mathcal{V}_i$ induced by the corresponding sequence of Last-Branch Recursion Paths. As demonstrated later in Chapter 6, the terminal set $\mathcal{V}_{i^*}$ coincides with the optimal set $\mathcal{U}^*$ from Theorem 3.2.8, providing the key link to the final optimality proof.

Throughout this and the following chapter, the phrase "consider an `RDCA`$(p, \mathcal{U}, \mathcal{J})$ program" (or simply "an `RDCA`$(p, \mathcal{U}, \mathcal{J})$ program") refers to the `RDCA` program for an

arbitrary but fixed input $p, \mathcal{U}, \mathcal{J}$, meeting the requirements specified in Algorithm 1. This encompasses the collection of program variables, their values, and the algebraic relationships that govern them during execution. The same convention applies analogously to the phrase "an $\texttt{DCA}(p, \mathcal{U})$ program".

5.1. Notational Conventions

The $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program has two structural characteristics that necessitate additional notation for referring to the values of program variables: (i) the `do-while` loop (lines 4–15), in which the values of the variables $(T, \boldsymbol{x})$, $\mathcal{Y}$, $\mathcal{U}$, and $\mathcal{K}$ are updated at each iteration; and (ii) potential further recursive invocations of `RDCA`, each with different values of the input parameters.

In the analysis that follows, we need to distinguish between the values of variables both across recursive invocations and across successive iterations of `do-while` loop within a given invocation of the program. To facilitate this, we introduce a notation that makes these distinctions explicit.

Recursive-invocation index. Consider the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program. To distinguish between the different recursive invocations triggered by this program, we introduce a recursion index $\boldsymbol{r}$ and denote by

$$\texttt{RDCA}(p, \mathcal{U}^{(\boldsymbol{r})}, \mathcal{J}^{(\boldsymbol{r})}) \tag{5.1}$$

the invocation associated with $\boldsymbol{r}$. The precise meaning and admissible values of $\boldsymbol{r}$ will be clarified in Section 5.4, where the *Last-Branch Recursion Path* is introduced. At this stage, it suffices to state that $\boldsymbol{r}$ will be defined as a pair of nonnegative integers.

To refer to the value of a variable in an invocation associated with $\boldsymbol{r}$, we attach the recursion index $\boldsymbol{r}$ to the variable. That is, every program variable in the invocation $\texttt{RDCA}(p, \mathcal{U}^{(\boldsymbol{r})}, \mathcal{J}^{(\boldsymbol{r})})$ carries this index. For example, $j^{(\boldsymbol{r})}$ denotes the value of the variable j in the $\texttt{RDCA}(p, \mathcal{U}^{(\boldsymbol{r})}, \mathcal{J}^{(\boldsymbol{r})})$ program.

For the outermost invocation $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$, this index is typically omitted. Nevertheless, in some cases we retain the index $\boldsymbol{r}$ even for the outermost invocation. This allows for a unified analysis of all invocations of interest, without the need to separately distinguish between the outermost and recursive invocations.

Loop-iteration index. To distinguish between different iterations of `do-while` loop within a given invocation $\texttt{RDCA}(p, \mathcal{U}^{(\boldsymbol{r})}, \mathcal{J}^{(\boldsymbol{r})})$, we introduce a second index i, defined as follows. For any variable z of the $\texttt{RDCA}(p, \mathcal{U}^{(\boldsymbol{r})}, \mathcal{J}^{(\boldsymbol{r})})$ program that is updated during `do-while` loop, we denote by

$$z^{(\boldsymbol{r};i)} \tag{5.2}$$

the value of z at the end of the i-th iteration of the loop, that is, immediately after line 14 and before line 15. The variables indexed in this manner include $(T, \boldsymbol{x})$, $\mathcal{Y}$, $\mathcal{U}$, and $\mathcal{K}$. The iteration index i ranges over the set I defined by

$$I := \begin{cases} \mathbb{N}, & \text{if } \; i^*_{\boldsymbol{r}} = \infty \\ \{i \in \mathbb{N}\colon \; i \leq i^*_{\boldsymbol{r}}\}, & \text{if } \; i^*_{\boldsymbol{r}} < \infty, \end{cases} \tag{5.3a}$$

where

$$i^*_{\boldsymbol{r}} := \inf\{i \in \mathbb{N}\colon \; \mathcal{Y}^{(\boldsymbol{r};i)} = \emptyset\}, \tag{5.3b}$$

denotes the final iteration of the loop in the invocation indexed by $\boldsymbol{r}$. For completeness, we adopt the convention $\inf \emptyset = \infty$, since at this stage we have not yet established that the set under the infimum in (5.3b) is nonempty. Proposition 5.2.2 later shows that $i^*_{\boldsymbol{r}} < \infty$.

Among the variables that are updated during `do-while` loop, two of them – $\mathcal{U}$ and $\mathcal{K}$ – already exist before the loop is entered. For these variables we define

$$\mathcal{U}^{(\boldsymbol{r};0)} := \mathcal{U}^{(\boldsymbol{r})} \qquad \text{and} \qquad \mathcal{K}^{(\boldsymbol{r};0)} := \mathcal{K}^{(\boldsymbol{r})}, \tag{5.4}$$

and we let the index $i = 0$ represent their state immediately before entering the loop. Only these two variables have well-defined values at iteration 0.

As noted in the previous paragraph, for the outermost invocation, we typically omit the recursion index $\boldsymbol{r}$. This also applies to the final loop iteration index $i^*_{\boldsymbol{r}}$ (written simply as i^*), as well as to loop-updated variables $z^{(\boldsymbol{r};i)}$ (written simply $z^{(i)}$), for $i \in \{1, \ldots, i^*\}$. In general, the applicable notation should be clear from context; whenever ambiguity might arise, it is stated explicitly.

In accordance with the indexing rules introduced in this section, we present an alternative formulation of the `RDCA` program using indexed variables (see Algorithm 3). In this version, all program variables are explicitly indexed by the recursion index r and the loop iteration i. This formulation provides a clearer view of the algorithm's progression and facilitates the subsequent analysis.

Algorithm 3 Indexed version of the `RDCA` algorithm

```
function RDCA(p, U^(r), J^(r))
    j^(r) ∈ J^(r)
    K^(r;0) ← {(d,h) ∈ H : d = j^(r)},  U^(r;0) ← U^(r)
    for i ∈ {1, ...} do
        if |J^(r)| = 1 then
            (T^(r;i), x^(r;i)) ← DCA(p, U^(r;i−1))
        else
            (T^(r;i), x^(r;i)) ← RDCA(p, U^(r;i−1), J^(r) \ {j^(r)})
        end if
        Y^(r;i) ← {(d,h) ∈ K^(r;i−1) : x^(r;i)_{d,h} ≥ N_{d,h}}        ▷ x^(r;i) = (x^(r;i)_{d,h}, (d,h) ∈ H)
        if Y^(r;i) ≠ ∅ then
            U^(r;i) ← U^(r;i−1) ∪ Y^(r;i)
            K^(r;i) ← K^(r;i−1) \ Y^(r;i)
        else
            i*_r ← i
            break
        end if
    end for
    return (T^(r;i*_r), x^(r;i*_r))
end function
```

To conclude this section, we remark that the vast majority of results in the subsequent sections, as well as in Chapter 6, concern only the outermost invocation $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$. For this reason, the recursion index $\boldsymbol{r}$ is typically omitted in accordance with the conventions established earlier. The use of recursive indices is reserved for specific cases where the recursion depth is analytically relevant (e.g., in Def. 5.5.1 and Lem. 5.4.3).

5.2. Definiteness and Termination

In this section, we prove that the `RDCA` program is well-defined and guaranteed to terminate. By well-definedness, we mean that the preconditions of both `RDCA` and `DCA` are satisfied whenever these programs are invoked. This property is formally stated in Proposition 5.2.2, which constitutes the main result of this section.

Lemma 5.2.1. *The* $\texttt{DCA}(p, \mathcal{U})$ *program returns* $(T, \boldsymbol{x})$ *satisfying:*

$$\boldsymbol{x} = (x_{d,h},\ (d,h) \in \mathcal{H}) \in \mathbb{R}_+^{|\mathcal{H}|}, \tag{5.5a}$$

$$x_{d,h} = N_{d,h}, \qquad (d,h) \in \mathcal{U}, \tag{5.5b}$$

$$\sum_{(d,h)\in\mathcal{H}} x_{d,h} = n, \tag{5.5c}$$

$$\sum_{h\in\eta_d(\mathcal{H})} \Big(\frac{1}{x_{d,h}} - \frac{1}{N_{d,h}}\Big)[A_{d,h}(p)]^2 = T, \qquad d \in \mathcal{D} := \begin{cases} \delta(\mathcal{H}), & \text{if } \mathcal{U} = \mathcal{H} \\ \delta(\mathcal{H}\setminus\mathcal{U}), & \text{if } \mathcal{U} \subsetneq \mathcal{H}. \end{cases} \tag{5.5d}$$

Remark 5.2.1. Based on line 8 of the DCA algorithm, the set $\mathcal{U}$ is interpreted as the collection of strata $(d,h) \in \mathcal{H}$ that are *blocked* at their upper bounds $N_{d,h}$. Consequently, following Remark 2.3.1 with $\mathcal{B} = \mathcal{U}$, the set $\delta(\mathcal{H}\setminus\mathcal{U})$ represents the domains that remain *unblocked* with respect to $\mathcal{U}$. When $\mathcal{U} \subsetneq \mathcal{H}$, the equation in (5.5d) is required to hold only for these unblocked domains; it may not hold for domains that are blocked by $\mathcal{U}$.

Proof of Lemma 5.2.1. Let $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$ and $\mathcal{U} \subseteq \mathcal{H}$ satisfy the precondition of $\texttt{DCA}(p, \mathcal{U})$, namely

$$\Big(n > \sum_{(d,h)\in\mathcal{U}} N_{d,h}\Big) \ \vee\ \Big(\mathcal{U} = \mathcal{H} \ \wedge\ n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big). \tag{5.6}$$

Case 1: $\mathcal{U} = \mathcal{H}$. By the definition of the algorithm, $T = 0$ and $x_{d,h} = N_{d,h}$ for all $(d,h) \in \mathcal{H}$. Then properties (5.5) hold immediately.

Case 2: $\mathcal{U} \subsetneq \mathcal{H}$. Define $\widetilde{p} = (\widetilde{\mathcal{H}}, \widetilde{\boldsymbol{N}}, \widetilde{\boldsymbol{S}}, \widetilde{\boldsymbol{\rho}}, \widetilde{n})$ with

$$\begin{aligned} &\widetilde{\mathcal{H}} := \mathcal{H}\setminus\mathcal{U}, \\ &\widetilde{\boldsymbol{N}} := (N_{d,h},\ (d,h) \in \widetilde{\mathcal{H}}), && \widetilde{\boldsymbol{\rho}} := (\rho_d,\ d \in \delta(\widetilde{\mathcal{H}})), \\ &\widetilde{\boldsymbol{S}} := (S_{d,h},\ (d,h) \in \widetilde{\mathcal{H}}), && \widetilde{n} := n - \sum_{(d,h)\in\mathcal{U}} N_{d,h}. \end{aligned} \tag{5.7}$$

Since $n - \sum_{(d,h)\in\mathcal{U}} N_{d,h} \overset{(5.6)}{>} 0$, the quintuple $\widetilde{p}$ satisfies Definition 2.2.2, and hence $\widetilde{p} \in \mathcal{P}$.

Let $(T, \boldsymbol{x})$ and $(\widetilde{T}, \widetilde{\boldsymbol{x}})$ be the points returned by $\texttt{DCA}(p, \mathcal{U})$ and $\texttt{DCA}(\widetilde{p}, \emptyset)$, respectively. By Definition 3.2.1, we have $\mathbf{D}_{\mathcal{U}|p} = \mathbf{D}_{\emptyset|\widetilde{p}}$, and consequently:

$$(\lambda, \boldsymbol{v}) := \mathrm{Eigen}(\mathcal{U} \mid p) = \mathrm{Eigen}(\emptyset \mid \widetilde{p}), \tag{5.8a}$$

$$\begin{aligned} s(\mathcal{U}, \boldsymbol{v}, d \mid p)\, A_{d,h}(p) &= \frac{n - \sum_{(i,j)\in\mathcal{U}} N_{i,j}}{\sum_{(i,j)\in\mathcal{H}\setminus\mathcal{U}} v_i A_{i,j}(p)}\, v_d\, A_{d,h}(p) \\ &= s(\emptyset, \boldsymbol{v}, d \mid \widetilde{p})\, A_{d,h}(\widetilde{p}), \qquad (d,h) \in \mathcal{H}\setminus\mathcal{U} = \widetilde{\mathcal{H}}. \end{aligned} \tag{5.8b}$$

From (5.8) and the definition of `DCA`, we obtain:

$$T = \widetilde{T} \qquad \text{and} \qquad x_{d,h} = \begin{cases} N_{d,h}, & (d,h) \in \mathcal{U} \\ \widetilde{x}_{d,h}, & (d,h) \in \mathcal{H} \setminus \mathcal{U} = \widetilde{\mathcal{H}}. \end{cases} \tag{5.9}$$

On the other hand, by Remark G.3, we have

$$\widetilde{\boldsymbol{x}} \in \mathbb{R}_+^{|\widetilde{\mathcal{H}}|}, \qquad \sum_{(d,h)\in\widetilde{\mathcal{H}}} \widetilde{x}_{d,h} = \widetilde{n}, \qquad \sum_{h\in\eta_d(\widetilde{\mathcal{H}})} \Big(\frac{1}{\widetilde{x}_{d,h}} - \frac{1}{N_{d,h}}\Big)[A_{d,h}(\widetilde{p})]^2 = \widetilde{T}, \;\; d \in \delta(\widetilde{\mathcal{H}}). \tag{5.10}$$

In view of (5.9) and (5.10), properties (5.5a) and (5.5b) hold directly. It remains to verify properties (5.5c)–(5.5d).

<u>(5.5c):</u>

$$\sum_{(d,h)\in\mathcal{H}} x_{d,h} \overset{(5.9)}{=} \sum_{(d,h)\in\mathcal{U}} N_{d,h} + \sum_{(d,h)\in\widetilde{\mathcal{H}}} \widetilde{x}_{d,h} \overset{(5.10)}{=} \sum_{(d,h)\in\mathcal{U}} N_{d,h} + \widetilde{n} \overset{(5.7)}{=} n. \tag{5.11}$$

<u>(5.5d):</u> For any $d \in \delta(\mathcal{H} \setminus \mathcal{U}) = \delta(\widetilde{\mathcal{H}})$, we have

$$\begin{aligned} \sum_{h\in\eta_d(\mathcal{H})} \Big(\frac{1}{x_{d,h}} - \frac{1}{N_{d,h}}\Big)[A_{d,h}(p)]^2 &\overset{(5.9)}{=} \sum_{h\in\eta_d(\widetilde{\mathcal{H}})} \Big(\frac{1}{\widetilde{x}_{d,h}} - \frac{1}{N_{d,h}}\Big)[A_{d,h}(\widetilde{p})]^2 \\ &\quad + \sum_{h\in\eta_d(\mathcal{U})} \Big(\frac{1}{N_{d,h}} - \frac{1}{N_{d,h}}\Big)[A_{d,h}(p)]^2 \\ &\overset{(5.10)}{=} \widetilde{T} \overset{(5.9)}{=} T. \end{aligned} \tag{5.12}$$

□

Proposition 5.2.2 (Definiteness and termination)**.** *Consider the* `RDCA`$(p, \mathcal{U}, \mathcal{J})$ *program and define:*

$$i^* := \inf\{i \in \mathbb{N}\colon\; \mathcal{Y}^{(i)} = \emptyset\} \qquad \textit{and} \qquad I := \begin{cases} \mathbb{N}, & \textit{if } i^* = \infty \\ \{i \in \mathbb{N}\colon\; i \le i^*\}, & \textit{if } i^* < \infty. \end{cases} \tag{5.13}$$

For each $i \in I$, the following hold:

$$\Big(\mathcal{U}^{(i-1)} \subseteq \mathcal{H}\Big) \;\wedge\; \Big[\Big(n > \sum_{(d,h)\in\mathcal{U}^{(i-1)}} N_{d,h}\Big) \;\vee\; \Big(\mathcal{U}^{(i-1)} = \mathcal{H} \;\wedge\; n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big)\Big], \tag{5.14a}$$

$$|\mathcal{J}| \ge 2 \quad\Longrightarrow\quad \Big[\Big(\mathcal{J} \setminus \{j\} \subseteq \delta(\mathcal{H})\Big) \;\wedge\; \Big(\delta(\mathcal{U}^{(i-1)}) \cap (\mathcal{J} \setminus \{j\}) = \emptyset\Big)\Big]. \tag{5.14b}$$

Moreover,

$$i^* < \infty, \tag{5.15}$$

and the point $(T, \boldsymbol{x})$ *returned by the program satisfies:*

$$\boldsymbol{x} = (x_{d,h},\ (d,h) \in \mathcal{H}) \in \mathbb{R}_+^{|\mathcal{H}|}, \tag{5.16a}$$

$$x_{d,h} = N_{d,h}, \qquad (d,h) \in \mathcal{U}, \tag{5.16b}$$

$$\sum_{(d,h)\in\mathcal{H}} x_{d,h} = n. \tag{5.16c}$$

Proof. Define the following predicate:

$B(k)$: For the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program with arbitrary valid input parameters such that $|\mathcal{J}| = k$, and with i^* and I defined in (5.13), properties (5.14)–(5.16) hold.

To prove the proposition, it suffices to show that $B(k)$ holds for all $k \in \mathbb{N}$, which we establish by induction. Whenever we refer to the context of $B(k)$, we mean the setting specified in its definition.

Base Case. We prove that $B(1)$ holds. All variables of the RDCA program, as well as i^* and I, are understood in the context of $B(1)$.

(5.14): For convenience, define the following propositional functions:

$$Q_{\mathcal{U}}(i) := \Big(\mathcal{U}^{(i)} \subseteq \mathcal{H}\Big) \wedge \Big[\Big(n > \sum_{(d,h)\in\mathcal{U}^{(i)}} N_{d,h}\Big) \vee \Big(\mathcal{U}^{(i)} = \mathcal{H} \ \wedge\ n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big)\Big],$$

$$i \in I \cup \{0\},$$

$$Q_x(i) := \Big(\boldsymbol{x}^{(i)} \in \mathbb{R}_+^{|\mathcal{H}|}\Big) \wedge \Big(\forall (d,h) \in \mathcal{U}^{(i-1)} \quad x_{d,h}^{(i)} = N_{d,h}\Big) \wedge \Big(\sum_{(d,h)\in\mathcal{H}} x_{d,h}^{(i)} = n\Big),$$

$$i \in I.$$

Let $i \in I$. Since $|\mathcal{J}| = 1$, the test on line 5 succeeds, and $\texttt{DCA}(p, \mathcal{U}^{(i-1)})$ is invoked on line 6 to obtain $(T^{(i)}, \boldsymbol{x}^{(i)})$. Thus, the following implications hold:

$$Q_{\mathcal{U}}(i-1) \quad \overset{\text{Lem. 5.2.1}}{\Longrightarrow} \quad Q_x(i) \quad \overset{[1]}{\Longrightarrow} \quad Q_{\mathcal{U}}(i), \tag{5.17}$$

where [1] is Lemma 2.2.1 applied with $\mathcal{A} = \mathcal{U}^{(i)}$ and $\boldsymbol{z} = \boldsymbol{x}^{(i)}$, noting that

$$\begin{aligned} \mathcal{U}^{(i)} &\overset{\text{line 12}}{=} \mathcal{U}^{(i-1)} \cup \mathcal{Y}^{(i)} \subseteq \mathcal{H}, \\ x_{d,h}^{(i)} &\overset{\text{line 10}}{\geq} N_{d,h}, \qquad (d,h) \in \mathcal{Y}^{(i)}. \end{aligned} \tag{5.18}$$

Since $Q_{\mathcal{U}}(0)$ holds as a precondition of the RDCA, it follows by transitivity of implication that $Q_{\mathcal{U}}(i)$ holds for all $i \in \{0\} \cup I$. Hence, (5.14a) is established for the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program with $|\mathcal{J}| = 1$.

The assertion (5.14b) is trivially satisfied when $|\mathcal{J}| = 1$.

(5.15): Define $S := \{i \in \mathbb{N}\colon\ \mathcal{Y}^{(i)} = \emptyset\}$ and let $\inf \emptyset := \infty$. According to line 13 of the program, for every $i \in \mathbb{N}$ such that $i < \inf S$, we have $\mathcal{K}^{(i-1)} \supsetneq \mathcal{K}^{(i)}$, which implies

$$|\mathcal{K}^{(i-1)}| > |\mathcal{K}^{(i)}|. \tag{5.19}$$

Suppose, for contradiction, that $S = \emptyset$. Then, by (5.19), the sequence $(|\mathcal{K}^{(i)}|)_{i\in\mathbb{N}_0}$ is strictly decreasing. However, a strictly decreasing sequence of nonnegative integers cannot be infinite. This is a contradiction. Hence $S \neq \emptyset$, and since $i^* := \inf S$, we conclude that $i^* < \infty$.

(5.16): Since (5.14)–(5.15) have been established, (5.16) follow from the first implication in (5.17) for $i = i^*$. In particular, (5.16b) holds in view of the inclusion $\mathcal{U} = \mathcal{U}^{(0)} \overset{\text{line 12}}{\subseteq} \mathcal{U}^{(i^*-1)}$.

Inductive Step. We aim to prove that for all $k \in \mathbb{N} \setminus \{1\}$, the implication $B(k-1) \Rightarrow B(k)$ holds. Let $k \in \mathbb{N} \setminus \{1\}$.

(5.14): In the context of $B(k)$, define the following propositional functions:

$$\begin{aligned}
Q_{\mathcal{U}}(i) &:= \Big(\mathcal{U}^{(i)} \subseteq \mathcal{H}\Big) \wedge \Big[\Big(n > \sum_{(d,h)\in\mathcal{U}^{(i)}} N_{d,h}\Big) \vee \Big(\mathcal{U}^{(i)} = \mathcal{H} \ \wedge\ n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big)\Big], \\
&\qquad i \in I \cup \{0\}, \\
Q_{\mathcal{J}}(i) &:= \Big[\Big(\mathcal{J} \setminus \{j\} \subseteq \delta(\mathcal{H})\Big) \wedge \Big(\delta(\mathcal{U}^{(i)}) \cap (\mathcal{J} \setminus \{j\}) = \emptyset\Big)\Big], \qquad i \in I \cup \{0\}, \\
Q_x(i) &:= \Big(\boldsymbol{x}^{(i)} \in \mathbb{R}_+^{|\mathcal{H}|}\Big) \wedge \Big(\forall (d,h) \in \mathcal{U}^{(i-1)} \quad x^{(i)}_{d,h} = N_{d,h}\Big) \wedge \Big(\sum_{(d,h)\in\mathcal{H}} x^{(i)}_{d,h} = n\Big), \\
&\qquad i \in I.
\end{aligned}$$

Let $i \in I$. Since $|\mathcal{J}| = k \geq 2$, the test on line 5 fails and $\texttt{RDCA}(p, \mathcal{U}^{(i-1)}, \mathcal{J} \setminus \{j\})$ is invoked on line 8 to obtain $(T^{(i)}, \boldsymbol{x}^{(i)})$. Thus, the following implications hold:

$$\big(Q_{\mathcal{U}}(i-1) \ \wedge\ Q_{\mathcal{J}}(i-1)\big) \quad \overset{[1]}{\Longrightarrow} \quad Q_x(i) \quad \overset{[2]}{\Longrightarrow} \quad Q_{\mathcal{U}}(i), \tag{5.20}$$

where:

[1]: The inductive hypothesis $B(k-1)$, specifically properties (5.15) and (5.16).

[2]: Lemma 2.2.1 applied with $\mathcal{A} = \mathcal{U}^{(i)}$ and $\boldsymbol{z} = \boldsymbol{x}^{(i)}$, noting that

$$\begin{aligned}
\mathcal{U}^{(i)} &\overset{\text{line 12}}{=} \mathcal{U}^{(i-1)} \cup \mathcal{Y}^{(i)} \subseteq \mathcal{H}, \\
x^{(i)}_{d,h} &\overset{\text{line 10}}{\geq} N_{d,h}, \qquad (d,h) \in \mathcal{Y}^{(i)}.
\end{aligned} \tag{5.21}$$

Furthermore,

$$Q_{\mathcal{J}}(i-1) \quad \Longrightarrow \quad Q_{\mathcal{J}}(i), \tag{5.22}$$

because

$$\delta(\mathcal{U}^{(i)}) \overset{\text{line 12}}{=} \delta(\mathcal{U}^{(i-1)} \cup \mathcal{Y}^{(i)}) = \delta(\mathcal{U}^{(i-1)}) \cup \delta(\mathcal{Y}^{(i)}) \overset{\substack{\text{line 3}\\ \text{line 10}}}{\subseteq} \delta(\mathcal{U}^{(i-1)}) \cup \{j\}. \tag{5.23}$$

Combining (5.20) with (5.22), we get:

$$\big(Q_{\mathcal{U}}(i-1) \ \wedge \ Q_{\mathcal{J}}(i-1)\big) \quad \Longrightarrow \quad \big(Q_{\mathcal{U}}(i) \ \wedge \ Q_{\mathcal{J}}(i)\big). \tag{5.24}$$

Since $Q_{\mathcal{U}}(0) \wedge Q_{\mathcal{J}}(0)$ holds as a precondition of the `RDCA`, it follows by transitivity of implication that $Q_{\mathcal{U}}(i) \wedge Q_{\mathcal{J}}(i)$ holds for all $i \in \{0\} \cup I$. Hence, (5.14) are established for the `RDCA`$(p, \mathcal{U}, \mathcal{J})$ program with $|\mathcal{J}| = k$.

(5.15): The proof of (5.15) in the context of $B(k)$ is identical to that of the base case $B(1)$, assuming the inductive hypothesis – specifically (5.15) – holds in the context of $B(k-1)$.

(5.16): Since (5.14)–(5.15) have been established, (5.16) follow from the first implication in (5.20) for $i = i^*$. In particular, (5.16b) holds in view of $\mathcal{U} = \mathcal{U}^{(0)} \overset{\text{line 12}}{\subseteq} \mathcal{U}^{(i^*-1)}$.

Since both the base case and the inductive step have been established, it follows by the principle of mathematical induction that $B(k)$ holds for every $k \in \mathbb{N}$. □

Remark 5.2.2. In the `RDCA`$(p, \mathcal{U}, \mathcal{J})$ program, a refined upper bound for i^* is:

$$i^* \leq |\gamma_j(\mathcal{H})| + 1. \tag{5.25}$$

Proof of Remark 5.2.2. According to lines 3 and 10, we have

$$\delta(\mathcal{Y}^{(i)}) \subseteq \{j\}, \qquad i \in \{1, \ldots, i^*\}. \tag{5.26}$$

Consequently,

$$\bigcup_{i=1}^{i^*-1} \mathcal{Y}^{(i)} \subseteq \gamma_j(\mathcal{H}). \tag{5.27}$$

Because the sets $\mathcal{Y}^{(i)}$, $i \in \{1, \ldots, i^*-1\}$, are nonempty (see line 15) and pairwise disjoint (see lines 10 and 13), the cardinality of the union is the sum of the cardinalities, with each term being at least 1. Thus,

$$i^* - 1 \leq \sum_{i=1}^{i^*-1} |\mathcal{Y}^{(i)}| = \Big|\bigcup_{i=1}^{i^*-1} \mathcal{Y}^{(i)}\Big| \overset{(5.27)}{\leq} |\gamma_j(\mathcal{H})|, \tag{5.28}$$

which yields (5.25). □

5.3. Basic Relations Between Program Variables

This section presents several fundamental relationships among the variables of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program. These simple identities follow directly from the algorithm's definition and provide the analytical foundation for the structural properties developed in this chapter, as well as for nearly all proofs presented in Chapter 6. For ease of reference, we collect them here in a single place.

By Proposition 5.2.2, the final iteration index i^* of the `do-while` loop is finite; hence, all subsequent expressions involving i^* are well-defined.

Corollary 5.3.1. *The variables of the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program satisfy:*

Properties of $\mathcal{Y}$*:*

$$\mathcal{Y}^{(i)} \neq \emptyset, \qquad i \in \{1, \ldots, i^* - 1\}, \qquad \text{[line 15]} \tag{5.29a}$$

$$\mathcal{Y}^{(i^*)} = \emptyset, \qquad \text{[line 15]} \tag{5.29b}$$

$$\mathcal{Y}^{(i_1)} \cap \mathcal{Y}^{(i_2)} = \emptyset, \qquad i_1, i_2 \in \{1, \ldots, i^*\},\ i_1 \neq i_2, \qquad \text{[lines 10, 13]} \tag{5.29c}$$

$$\delta(\mathcal{Y}^{(i)}) \subseteq \{j\}, \qquad i \in \{1, \ldots, i^*\}. \qquad \text{[lines 3, 10]} \tag{5.29d}$$

Properties of $\mathcal{U}$*:*

$$\gamma_{\mathcal{J}}(\mathcal{U}) = \emptyset, \qquad \text{[RDCA prereq.]} \tag{5.30a}$$

$$\mathcal{U}^{(i)} = \mathcal{U} \cup \bigcup_{t=1}^{i} \mathcal{Y}^{(t)}, \qquad i \in \{0, \ldots, i^*\}. \qquad \text{[see proof below]} \tag{5.30b}$$

Properties of $\mathcal{K}$*:*

$$\mathcal{K}^{(i)} = \gamma_j(\mathcal{H}) \setminus \bigcup_{t=1}^{i} \mathcal{Y}^{(t)}, \qquad i \in \{0, \ldots, i^*\}. \qquad \text{[see proof below]} \tag{5.31}$$

Proof. All statements follow directly from the algorithm's construction, with justifications given in parentheses. Two relations, however, are proved below.

<u>(5.30b):</u> By the notational conventions established in Section 5.1, $\mathcal{U}^{(0)} = \mathcal{U}$.
For $i \in \{1, \ldots, i^*\}$, line 12 and $\mathcal{Y}^{(i^*)} \overset{(5.29b)}{=} \emptyset$ together yield

$$\mathcal{U}^{(i)} = \mathcal{U}^{(i-1)} \cup \mathcal{Y}^{(i)}. \tag{5.32}$$

<u>(5.31):</u> The identity $\mathcal{K}^{(0)} = \gamma_j(\mathcal{H})$ follows from the initialization on line 3.
For $i \in \{1, \ldots, i^*\}$, line 13 and $\mathcal{Y}^{(i^*)} \overset{(5.29b)}{=} \emptyset$ together yield

$$\mathcal{K}^{(i)} = \mathcal{K}^{(i-1)} \setminus \mathcal{Y}^{(i)}. \tag{5.33}$$

□

5.4. The Last-Branch Recursion Path

In Section 5.1, we introduced a notational convention to refer to the values realized by program variables across recursive invocations of the `RDCA` program. At that stage, the recursion index $\boldsymbol{r}$ was defined only in principle, without specifying its formal structure or the set of admissible values it may assume. In the present section, we provide this clarification in full through the introduction of the Last-Branch Recursion Path.

This delayed specification is intentional: the formal construction of the Last-Branch Recursion Path in Definition 5.4.1 presupposes the well-definedness of the `RDCA` program and its termination (i.e., that $i^* < \infty$), results established in Section 5.2 (Prop. 5.2.2).

Definition 5.4.1 (Last-Branch Recursion Path)**.** Consider the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program, and let $i \in \{1, \ldots, i^*\}$ denote the iteration index of the `do-while` loop, as specified in Section 5.1.

The *Last-Branch Recursion Path rooted at* i, abbreviated $\mathrm{LBRP}(i \mid p, \mathcal{U}, \mathcal{J})$, is the sequence of invocations of `RDCA` defined by

$$\mathrm{LBRP}(i \mid p, \mathcal{U}, \mathcal{J}) := \left\{\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})\right\}_{r \in \{0, \ldots, |\mathcal{J}|-1\}}, \tag{5.34}$$

where:

- $\{0, \ldots, |\mathcal{J}|-1\}$ is the set of recursion levels of the $\mathrm{LBRP}(i \mid p, \mathcal{U}, \mathcal{J})$, where $|\mathcal{J}|-1$ is the maximum recursion depth reached before a base case.
- The index r denotes the current recursion level (i.e., the depth along the path).
- $\mathcal{U}^{(i,r)} := \begin{cases} \mathcal{U}, & \text{if } r = 0 \\ \mathcal{U}^{(i,r-1;\tilde{i}-1)}, & \text{if } r \geq 1, \end{cases} \qquad \tilde{i} := \begin{cases} i, & \text{if } r = 1 \\ i^*_{i,r-1}, & \text{if } r \geq 2, \end{cases}$

  where:
  - $\mathcal{U}^{(i,r-1;\tilde{i}-1)}$ denotes the value of $\mathcal{U}$ at the end of the $(\tilde{i}-1)$-th iteration of the `do-while` loop in the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r-1)}, \mathcal{J}^{(i,r-1)})$ at recursion level $r-1$ of the $\mathrm{LBRP}(i \mid p, \mathcal{U}, \mathcal{J})$;
  - $i^*_{i,r-1}$ denotes the final iteration of the `do-while` loop in the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r-1)}, \mathcal{J}^{(i,r-1)})$, and is defined as in (5.3b) for $\boldsymbol{r} = (i, r-1)$.
- $\mathcal{J}^{(i,r)} := \begin{cases} \mathcal{J}, & \text{if } r = 0 \\ \mathcal{J}^{(i,r-1)} \setminus \{j^{(i,r-1)}\}, & \text{if } r \geq 1, \end{cases}$

  where $j^{(i,r-1)}$ denotes the value of j (see line 2) in the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r-1)}, \mathcal{J}^{(i,r-1)})$ at level $r-1$ of the $\mathrm{LBRP}(i \mid p, \mathcal{U}, \mathcal{J})$.

Definition 5.4.1 formalizes the recursion index $\boldsymbol{r}$ introduced in Section 5.1. Specifically, for the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program and a given iteration $i \in \{1, \ldots, i^*\}$, the abstract index $\boldsymbol{r}$ is realized as the pair (i, r), which uniquely identifies the recursive invocation at level $r \in \{0, \ldots, |\mathcal{J}| - 1\}$ along the $\text{LBRP}(i \mid p, \mathcal{U}, \mathcal{J})$, namely $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$.

While the algorithm's execution generates a branching recursion tree, the LBRP identifies a single linear path within this structure (see Fig. 5.4.1, shown later). As the subsequent analysis will indicate, the invocations along this specific path are sufficient for establishing the correctness of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program with $\mathcal{U} = \emptyset$ and $\mathcal{J} = \delta(\mathcal{H})$.

Recursion level $r = 0$. The first element of an $\text{LBRP}(i \mid p, \mathcal{U}, \mathcal{J})$ is the outermost invocation $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$. Its inclusion in the sequence facilitates a unified treatment of the recursion, allowing this outermost invocation to be analyzed without being treated as a special case.

Consistent with the conventions in Section 5.1, the recursion index $(i, 0)$ for the outermost invocation is generally omitted. It is utilized primarily in expressions where program variables are explicitly indexed by r over $\{0, \ldots, |\mathcal{J}| - 1\}$, thereby subsuming the case $r = 0$ into the general formulation.

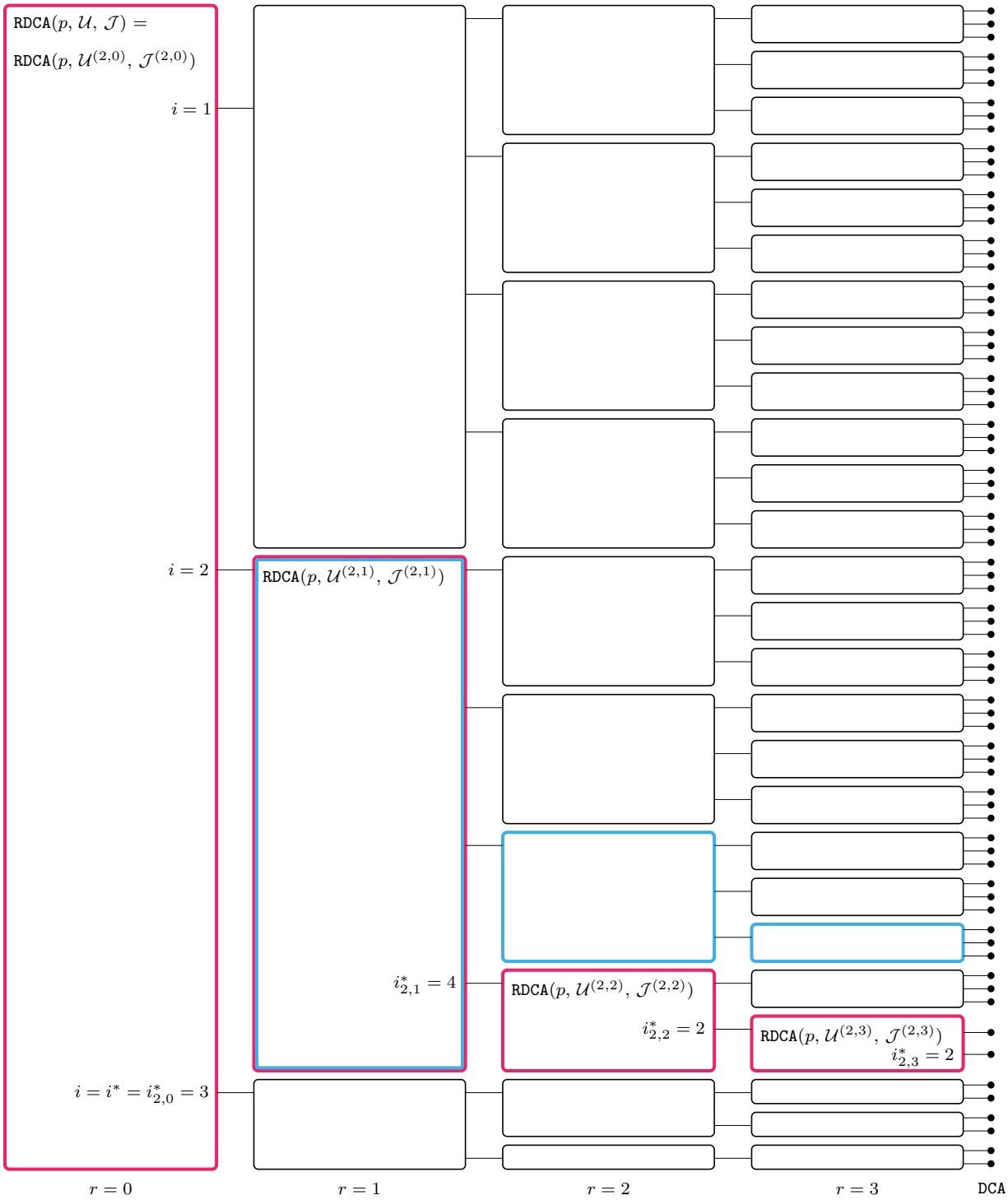
According to the conventions introduced in Section 5.1 and the definition of the $\text{LBRP}(i \mid p, \mathcal{U}, \mathcal{J})$, the following notational identities hold for any $i \in \{1, \ldots, i^*\}$:

$$\begin{aligned} y^{(i,0)} &= y, \\ z^{(i,0;t)} &= z^{(t)}, \qquad t \in \{1, \ldots, i^*_{i,0}\}, \\ i^*_{i,0} &= i^*, \end{aligned} \tag{5.35}$$

where y denotes the variables j and $\mathcal{J}$, while z denotes $(T, \boldsymbol{x})$, $\mathcal{Y}$, $\mathcal{U}$, $\mathcal{K}$.

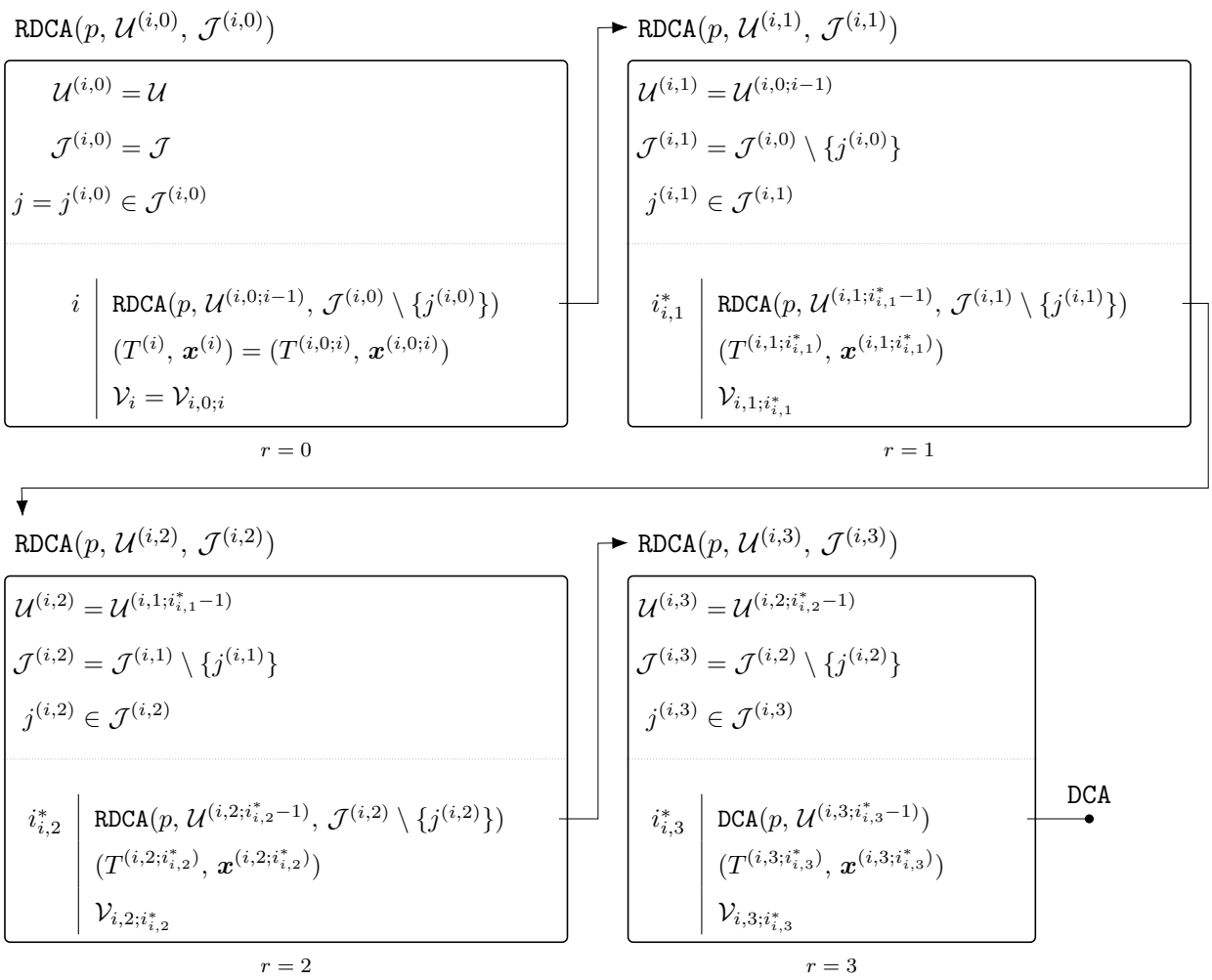
To aid understanding, Figure 5.4.1 presents a schematic recursion tree of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ algorithm for certain input parameters. The figure highlights two specific Last-Branch Recursion Paths: $\text{LBRP}(2 \mid p, \mathcal{U}, \mathcal{J})$ and the subsequent $\text{LBRP}(3 \mid p, \mathcal{U}^{(2,1)}, \mathcal{J}^{(2,1)})$, the latter of which originates from a recursive invocation within the former. Although the formal details of this secondary recursion path will be expanded upon in the sequel, it is included here to provide a comprehensive visual representation and to preclude the need for a separate diagram in later sections.

Figure 5.4.1: Example recursion tree of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program with $|\mathcal{J}| = 4$, showing two Last-Branch Recursion Paths: $\text{LBRP}(2 \mid p, \mathcal{U}, \mathcal{J})$ (red) and $\text{LBRP}(3 \mid p, \mathcal{U}^{(2,1)}, \mathcal{J}^{(2,1)})$ (blue). Each box represents a (recursive) invocation of RDCA.



To illustrate Definition 5.4.1, Figure 5.4.2 provides a detailed depiction of the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$) for an arbitrary $i \in \{1, \ldots, i^*\}$ and $|\mathcal{J}| = 4$. The figure traces the sequence of recursive invocations along the path, illustrating how each invocation triggers the next until terminating in the base-case invocation `DCA`. Furthermore, the figure displays the notation for the program variable $(T, \boldsymbol{x})$ and the sets $\mathcal{V}_{\cdot}$ (to be formally introduced in Section 5.5) within the context of the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$). Consolidating these elements here avoids redundant diagrams and highlights the structural patterns that characterize a Last-Branch Recursion Path.

Figure 5.4.2: Illustration of the Last-Branch Recursion Path for the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program with $|\mathcal{J}| = 4$, rooted at $i \in \{1, \ldots, i^*\}$. Each box represents a (recursive) invocation of `RDCA` along this path, while arrows indicate the transitions between successive invocations. The lower part of each box shows the call that initiates the next (recursive) invocation, along with the notation for $(T, \boldsymbol{x})$ and the sets $\mathcal{V}_{\cdot}$ (defined later in Sec. 5.5).



As illustrated in Figure 5.4.2, each box represents a recursive `RDCA` invocation along the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$). For the case $|\mathcal{J}| = 4$, the path comprises four recursion levels: $\{0, 1, 2, 3\}$. The first box ($r = 0$) signifies the outermost invocation, and the subsequent invocation along the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$) is

$$\texttt{RDCA}(p, \mathcal{U}^{(i,0;i-1)}, \mathcal{J}^{(i,0)} \setminus \{j^{(i,0)}\}) = \texttt{RDCA}(p, \mathcal{U}^{(i,1)}, \mathcal{J}^{(i,1)}), \tag{5.36}$$

which is triggered during the i-th iteration of the `do-while` loop.

For recursion levels $r \geq 1$, the invocations along the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$) follow the final iterations of the respective `do-while` loops. For $r = 1$, this yields

$$\texttt{RDCA}(p, \mathcal{U}^{(i,1;i^*_{i,1}-1)}, \mathcal{J}^{(i,1)} \setminus \{j^{(i,1)}\}) = \texttt{RDCA}(p, \mathcal{U}^{(i,2)}, \mathcal{J}^{(i,2)}), \tag{5.37}$$

with a similar transition occurring for $r = 2$. The path reaches its maximum recursion depth at $r = 3$, at which point, $|\mathcal{J}^{(i,3)}| = 1$, leading to the base-case invocation.

Secondary LBRPs **arising from the** LBRP($i \mid p, \mathcal{U}, \mathcal{J}$)**.** According to Definition 5.4.1, an LBRP($i \mid p, \mathcal{U}, \mathcal{J}$) can originate from an `RDCA`($p, \mathcal{U}, \mathcal{J}$) program for any admissible input parameters. In Section 5.5, it will be necessary to consider a secondary LBRP originating from an arbitrary recursive invocation along the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$). An example of such a path, namely the LBRP($3 \mid p, \mathcal{U}^{(2,1)}, \mathcal{J}^{(2,1)}$), was already illustrated in Figure 5.4.1 prior to its formal specification. The following remark describes how Definition 5.4.1 is used to construct such a subsequent path.

Remark 5.4.1 (Secondary LBRPs arising from the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$))**.** Consider the `RDCA`($p, \mathcal{U}, \mathcal{J}$) program and let $i \in \{1, \ldots, i^*\}$ be an arbitrary iteration with the corresponding LBRP($i \mid p, \mathcal{U}, \mathcal{J}$). Let $r \in \{0, \ldots, |\mathcal{J}| - 1\}$. Applying Definition 5.4.1 to the invocation `RDCA`($p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)}$) along the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$), yields the secondary path

$$\text{LBRP}(g \mid p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)}), \tag{5.38}$$

where the iteration index g of the `do-while` loop in this invocation ranges over $\{1, \ldots, i^*_{i,r}\}$.

Remark 5.4.2. A fully explicit formula for the secondary path LBRP($g \mid p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)}$) in Remark 5.4.1 generally requires an extended indexing scheme to account for `RDCA` invocations that lie outside the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$). However, in the two cases relevant

to our analysis – namely $r = 0$ and $g = i^*_{i,r}$ for $r \geq 1$ – the LBRP$(g \mid p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ is a subpath of the LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$, allowing the existing indexing to suffice. This structural simplification yields:

$$\text{LBRP}(g \mid p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)}) = \left\{\texttt{RDCA}(p, \mathcal{U}^{(i,\ell)}, \mathcal{J}^{(i,\ell)})\right\}_{\ell \in \{r, \ldots, |\mathcal{J}|-1\}}, \tag{5.39a}$$

for

$$g = \begin{cases} i, & \text{if } r = 0 \\ i^*_{i,r}, & \text{if } r \geq 1. \end{cases} \tag{5.39b}$$

Lemma 5.4.1. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program and let* $i \in \{1, \ldots, i^*\}$ *be an arbitrary iteration with the corresponding* LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$*. Then,*

$$\mathcal{J} = \bigcup_{r=0}^{|\mathcal{J}|-1} \{j^{(i,r)}\}. \tag{5.40}$$

Proof. Let $R := \{0, \ldots, |\mathcal{J}| - 1\}$. By construction of the LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$, at each level $r \in R$ along this path, we have

$$j^{(i,r)} \in \mathcal{J}^{(i,r)} = \begin{cases} \mathcal{J}, & \text{if } r = 0 \\ \mathcal{J} \setminus \bigcup_{\ell=0}^{r-1} \{j^{(i,\ell)}\}, & \text{if } r \geq 1. \end{cases} \tag{5.41}$$

Therefore, $\{j^{(i,r)} \colon r \in R\} \subset \mathcal{J}$.

Define the map $\phi \colon R \to \mathcal{J}$ by $\phi(r) := j^{(i,r)}$. From (5.41), for any $r_1, r_2 \in R$ with $r_1 < r_2$, we have

$$j^{(i,r_2)} \neq j^{(i,r_1)}. \tag{5.42}$$

If $r_2 < r_1$, the same argument yields $j^{(i,r_1)} \neq j^{(i,r_2)}$, which is equivalent to the above. Thus, $j^{(i,r_1)} \neq j^{(i,r_2)}$ whenever $r_1 \neq r_2$, and therefore ϕ is injective.

Since both R and $\mathcal{J}$ are finite sets with $|R| = |\mathcal{J}|$, it follows that ϕ is also surjective. Hence, ϕ is bijective, and the identity (5.40) follows. □

Lemma 5.4.2. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program and let* $i \in \{1, \ldots, i^*\}$ *be an arbitrary iteration with the corresponding* LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$*. Then,*

$$|\mathcal{J}^{(i,r)}| = |\mathcal{J}| - r, \qquad r \in \{0, \ldots, |\mathcal{J}| - 1\}. \tag{5.43}$$

Proof. The proof proceeds by induction on $r \in \{0, \ldots, |\mathcal{J}| - 1\}$.

Base Case. By Definition 5.4.1 of the LBRP, $\mathcal{J}^{(i,0)} = \mathcal{J}$, so $|\mathcal{J}^{(i,0)}| = |\mathcal{J}|$.

Inductive Step. Assume $|\mathcal{J}| \geq 2$ and fix $r \in \{1, \ldots, |\mathcal{J}| - 1\}$. As the inductive hypothesis, assume that

$$|\mathcal{J}^{(i,r-1)}| = |\mathcal{J}| - (r-1). \tag{5.44}$$

It then follows that:

$$|\mathcal{J}^{(i,r)}| \overset{\substack{\text{Def. 5.4.1} \\ r \geq 1}}{=} |\mathcal{J}^{(i,r-1)} \setminus \{j^{(i,r-1)}\}| \overset{[1]}{=} |\mathcal{J}^{(i,r-1)}| - 1 \overset{(5.44)}{=} |\mathcal{J}| - r, \tag{5.45}$$

where [1] holds because $j^{(i,r-1)} \in \mathcal{J}^{(i,r-1)}$ (see line 2).

As the choice of r was arbitrary, the inductive step is valid for all $r \in \{1, \ldots, |\mathcal{J}|-1\}$. By the principle of mathematical induction, (5.43) holds. □

Lemma 5.4.3. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program and let* $i \in \{1, \ldots, i^*\}$ *be an arbitrary iteration with the corresponding* $\text{LBRP}(i \mid p, \mathcal{U}, \mathcal{J})$. *Then,*

$$(T^{(i)}, \boldsymbol{x}^{(i)}) = (T^{(i,r;\tilde{i}_r)}, \boldsymbol{x}^{(i,r;\tilde{i}_r)}), \qquad r \in \{0, \ldots, |\mathcal{J}| - 1\}, \tag{5.46a}$$

where

$$\tilde{i}_r := \begin{cases} i, & \text{if } r = 0 \\ i^*_{i,r}, & \text{if } r \geq 1. \end{cases} \tag{5.46b}$$

Proof. For $r = 0$, the identity in (5.46a) holds trivially by the convention (5.35), namely:

$$(T^{(i)}, \boldsymbol{x}^{(i)}) = (T^{(i,0;i)}, \boldsymbol{x}^{(i,0;i)}). \tag{5.47}$$

To prove the lemma, it remains to establish the following sequence of identities:

$$(T^{(i,r;\tilde{i}_r)}, \boldsymbol{x}^{(i,r;\tilde{i}_r)}) = (T^{(i,r+1;i^*_{i,r+1})}, \boldsymbol{x}^{(i,r+1;i^*_{i,r+1})}), \qquad r \in \{0, \ldots, |\mathcal{J}| - 2\}. \tag{5.48}$$

Assume that $|\mathcal{J}| \geq 2$ and fix $r \in \{0, \ldots, |\mathcal{J}|-2\}$. To simplify the notation, let $\tilde{i} := \tilde{i}_r$.

Since $r < |\mathcal{J}| - 1$, by Lemma 5.4.2, $|\mathcal{J}^{(i,r)}| > 1$. Consequently, in the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ along the $\text{LBRP}(i \mid p, \mathcal{U}, \mathcal{J})$, the condition on line 5 is not satisfied. By the recursive call on line 8, we obtain:

$$\begin{aligned} (T^{(i,r;\tilde{i})}, \boldsymbol{x}^{(i,r;\tilde{i})}) &= \texttt{RDCA}(p, \mathcal{U}^{(i,r;\tilde{i}-1)}, \mathcal{J}^{(i,r)} \setminus \{j^{(i,r)}\}) \\ &\overset{[1]}{=} \texttt{RDCA}(p, \mathcal{U}^{(i,r+1)}, \mathcal{J}^{(i,r+1)}), \end{aligned} \tag{5.49}$$

where [1] follows directly from Definition 5.4.1 of the LBRP.

On the other hand, $\texttt{RDCA}(p, \mathcal{U}^{(i,r+1)}, \mathcal{J}^{(i,r+1)})$ returns the pair

$$(T^{(i,r+1;i^*_{i,r+1})}, \boldsymbol{x}^{(i,r+1;i^*_{i,r+1})}). \tag{5.50}$$

Combining (5.49) with (5.50) establishes (5.48) for the chosen r. Since $r \in \{0, \ldots, |\mathcal{J}| - 2\}$ was arbitrary, the identity in (5.48) holds for all r in this range. □

5.5. The take-max Strata Sets $\mathcal{V}_i$

In this section we define a sequence of sets $\mathcal{V}_i \subseteq \mathcal{H}$, $i \in \{1, \ldots, i^*\}$, which arise in connection with the Last-Branch Recursion Path introduced in Section 5.4. These sets play a central role in establishing the total correctness of $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ when $\mathcal{U} = \emptyset$ and $\mathcal{J} = \delta(\mathcal{H})$. Specifically, as shown in Section 6.4 of Chapter 6, when the program is executed with these parameter values, it identifies the set $\mathcal{V}_{i^*}$, which coincides with the set $\mathcal{U}^* \subseteq \mathcal{H}$ characterizing the optimal solution in Theorem 3.2.8. Since that theorem provides a closed-form expression for the optimal solution in terms of $\mathcal{U}^*$, identifying the sets $\mathcal{V}_i$ – and in particular $\mathcal{V}_{i^*}$ – is a crucial step in proving that the `RDCA` program solves the CPDA problem. After defining these sets, we establish several of their fundamental properties.

Definition 5.5.1 (The sets $\mathcal{V}_i$)**.** Consider the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program. For each iteration $i \in \{1, \ldots, i^*\}$ and the corresponding LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$, the set $\mathcal{V}_i$ is defined by

$$\mathcal{V}_i := \mathcal{U} \cup \bigcup_{t=1}^{i-1} \mathcal{Y}^{(t)} \cup \bigcup_{r=1}^{|\mathcal{J}|-1} \bigcup_{t=1}^{i^*_{i,r}-1} \mathcal{Y}^{(i,r;t)}, \tag{5.51}$$

where $\mathcal{Y}^{(i,r;t)}$ denotes the value of $\mathcal{Y}$ at the end of the t-th iteration of the `do-while` loop in the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ along the LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$.

The sets $\mathcal{V}_i$ relative to a recursive invocation. As established in Definition 5.5.1, the sets $\mathcal{V}_i$, $i \in \{1, \ldots, i^*\}$, are defined for the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program for any admissible input parameters. In the correctness analysis in Chapter 6, we sometimes need to consider the sets defined by (5.51), but relative to the recursive invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ along the LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$ for some $r \in \{1, \ldots, |\mathcal{J}| - 1\}$. The following two remarks provide the formal specification for the resulting sets.

Remark 5.5.1 (The sets $\mathcal{V}_i$ relative to a recursive invocation)**.** Consider the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program and fix an arbitrary iteration $i \in \{1, \ldots, i^*\}$ with the corresponding LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$. For an arbitrary $r \in \{0, \ldots, |\mathcal{J}| - 1\}$, consider the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ along this path. Applying Definition 5.5.1 to the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ at an arbitrary iteration $g \in \{1, \ldots, i^*_{i,r}\}$ of its loop, with the corresponding LBRP$(g \mid p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$, yields the set

$$\mathcal{V}_{i,r;g}. \tag{5.52}$$

Here, LBRP($g \mid p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)}$) is the secondary recursion path as described in Remark 5.4.1. The set $\mathcal{V}_{i,r;g}$ is the analogue of $\mathcal{V}_i$ for iteration g within the recursive invocation `RDCA`($p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)}$).

Remark 5.5.2. A fully explicit formula for the set $\mathcal{V}_{i,r;g}$ in Remark 5.5.1 generally requires an extended indexing scheme to account for `RDCA` invocations along the LBRP($g \mid p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)}$) that lie outside the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$). In our analysis, however, we only require the formula in the two special cases: $r = 0$ and $g = i^*_{i,r}$ for $r \geq 1$. In these cases, the notation simplifies significantly because the LBRP($g \mid p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)}$) is then a subpath of the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$) (see Rem. 5.4.2), allowing the existing indexing to suffice. This structural simplification yields:

$$\begin{aligned}\mathcal{V}_{i,r;\tilde{i}_r} &\overset{(5.51)}{=} \mathcal{U}^{(i,r)} \cup \bigcup_{t=1}^{\tilde{i}_r-1} \mathcal{Y}^{(i,r;t)} \cup \bigcup_{\ell=1}^{|\mathcal{J}^{(i,r)}|-1} \bigcup_{t=1}^{i^*_{i,r+\ell}-1} \mathcal{Y}^{(i,r+\ell;t)} \\ &\overset{(5.43)}{=} \mathcal{U}^{(i,r)} \cup \bigcup_{t=1}^{\tilde{i}_r-1} \mathcal{Y}^{(i,r;t)} \cup \bigcup_{\ell=r+1}^{|\mathcal{J}|-1} \bigcup_{t=1}^{i^*_{i,\ell}-1} \mathcal{Y}^{(i,\ell;t)},\end{aligned} \tag{5.53}$$

where $\tilde{i}_r := \begin{cases} i, & \text{if } r = 0 \\ i^*_{i,r}, & \text{if } r \geq 1. \end{cases}$

Following Remark 5.5.2, for $r = 0$, the identity in (5.53) reduces to (5.51) as expected:

$$\mathcal{V}_{i,0;i} = \mathcal{U}^{(i,0)} \cup \bigcup_{t=1}^{i-1} \mathcal{Y}^{(i,0;t)} \cup \bigcup_{\ell=1}^{|\mathcal{J}|-1} \bigcup_{t=1}^{i^*_{i,\ell}-1} \mathcal{Y}^{(i,\ell;t)} \overset{(5.35)}{=} \mathcal{V}_i. \tag{5.54}$$

According to Definition 5.5.1, the set $\mathcal{V}_i$ corresponds to the i-th iteration of the `do-while` loop in the `RDCA`($p, \mathcal{U}, \mathcal{J}$) program. Similarly, Remark 5.5.1 specifies that the set $\mathcal{V}_{i,r;g}$ corresponds to the g-th iteration of the `do-while` loop in the invocation `RDCA`($p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)}$) along the LBRP($i \mid p, \mathcal{U}, \mathcal{J}$). Figure 5.4.2 illustrates this relationship for $g = \tilde{i}_r$, consolidating the LBRP and the associated sets $\mathcal{V}_\cdot$ into a single diagram.

Standing convention for sets $\mathcal{V}_\cdot$. Throughout the remainder of this chapter and the entirety of the following chapter, the sets $\mathcal{V}_\cdot$ are to be understood as defined in Definition 5.5.1 or Remark 5.5.1, relative to the prevailing `RDCA`. Since the appropriate context is generally clear from the problem setting, we shall not explicitly reference these definitions or restate the context unless ambiguity arises, in which case it will be specified.

Lemma 5.5.1. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program and let* $i \in \{1, \ldots, i^*\}$. *Then,*

$$\mathcal{V}_i = \mathcal{V}_{i,r;\tilde{i}_r}, \qquad r \in \{0, \ldots, |\mathcal{J}| - 1\}, \tag{5.55a}$$

where the sets $\mathcal{V}_i$ *and* $\mathcal{V}_{i,r;\tilde{i}_r}$ *are defined as in* (5.51) *and* (5.53), *respectively, with*

$$\tilde{i}_r := \begin{cases} i, & \text{if } r = 0 \\ i^*_{i,r}, & \text{if } r \geq 1. \end{cases} \tag{5.55b}$$

Proof. Define the following predicate:

$B(k \mid i)$: The identity in (5.55) holds for $r = k$.

To prove the lemma, it suffices to show that $B(k \mid i)$ is true for all $k \in \{0, \ldots, |\mathcal{J}| - 1\}$. The proof proceeds by induction on k.

Base Case. By (5.54), we have $\mathcal{V}_i = \mathcal{V}_{i,0;i}$. Hence, $B(0 \mid i)$ holds trivially.

Inductive Step. We aim to prove that the implication $B(k-1 \mid i) \Rightarrow B(k \mid i)$ holds for all $k \in \{1, \ldots, |\mathcal{J}| - 1\}$.

Assume $|\mathcal{J}| \geq 2$ and fix $k \in \{1, \ldots, |\mathcal{J}| - 1\}$. Let $\tilde{i} := \begin{cases} i, & \text{if } k = 1 \\ i^*_{i,k-1}, & \text{if } k \geq 2. \end{cases}$

As the inductive hypothesis, assume that $B(k-1 \mid i)$ holds, i.e.,

$$\mathcal{V}_i = \mathcal{V}_{i,k-1;\tilde{i}}. \tag{5.56}$$

By Definition 5.4.1 of the LBRP, since $k \geq 1$,

$$\mathcal{U}^{(i,k)} = \mathcal{U}^{(i,k-1;\tilde{i}-1)} \overset{(5.30\text{b})}{=} \mathcal{U}^{(i,k-1)} \cup \bigcup_{t=1}^{\tilde{i}-1} \mathcal{Y}^{(i,k-1;t)}. \tag{5.57}$$

Furthermore,

$$\begin{aligned} \mathcal{V}_{i,k;i^*_{i,k}} &\overset{\substack{(5.53)\\ k\geq 1}}{=} \mathcal{U}^{(i,k)} \cup \bigcup_{\ell=k}^{|\mathcal{J}|-1} \bigcup_{t=1}^{i^*_{i,\ell}-1} \mathcal{Y}^{(i,\ell;t)} \\ &\overset{(5.57)}{=} \mathcal{U}^{(i,k-1)} \cup \bigcup_{t=1}^{\tilde{i}-1} \mathcal{Y}^{(i,k-1;t)} \cup \bigcup_{\ell=k}^{|\mathcal{J}|-1} \bigcup_{t=1}^{i^*_{i,\ell}-1} \mathcal{Y}^{(i,\ell;t)} \\ &\overset{(5.53)}{=} \mathcal{V}_{i,k-1;\tilde{i}}. \end{aligned} \tag{5.58}$$

Substituting (5.58) into (5.56) yields $\mathcal{V}_i = \mathcal{V}_{i,k;i^*_{i,k}}$, confirming $B(k \mid i)$. As the choice of k was arbitrary, the inductive step holds for all $k \in \{1, \ldots, |\mathcal{J}| - 1\}$.

By the principle of mathematical induction, $B(k \mid i)$ holds for all $k \in \{0, \ldots, |\mathcal{J}|-1\}$. □

Lemma 5.5.2. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program and let* $\mathcal{A} \subseteq \delta(\mathcal{H})$. *For any* $i \in \{1, \ldots, i^*\}$, *it holds that*

$$\gamma_{\mathcal{A}}(\mathcal{V}_i) = \begin{cases} \mathcal{U}, & \text{if } \mathcal{A} \subseteq \delta(\mathcal{H}) \setminus \mathcal{J} \\ \bigcup_{t=1}^{i-1} \mathcal{Y}^{(t)}, & \text{if } \mathcal{A} = \{j\}. \end{cases} \tag{5.59}$$

Proof. Let $i \in \{1, \ldots, i^*\}$. By definition (5.51) and in view of (2.17a), we have

$$\gamma_{\mathcal{A}}(\mathcal{V}_i) = \gamma_{\mathcal{A}}(\mathcal{U}) \cup \bigcup_{t=1}^{i-1} \gamma_{\mathcal{A}}(\mathcal{Y}^{(t)}) \cup \bigcup_{r=1}^{|\mathcal{J}|-1} \bigcup_{t=1}^{i^*_{i,r}-1} \gamma_{\mathcal{A}}(\mathcal{Y}^{(i,r;t)}). \tag{5.60}$$

For any $t \in \{1, \ldots, i-1\}$, the definition of the algorithm yields

$$\gamma_{\mathcal{A}}(\mathcal{U}) \overset{(5.30a)}{=} \begin{cases} \mathcal{U}, & \text{if } \mathcal{A} \subseteq \delta(\mathcal{H}) \setminus \mathcal{J} \\ \emptyset, & \text{if } \mathcal{A} = \{j\}, \end{cases} \qquad \gamma_{\mathcal{A}}(\mathcal{Y}^{(t)}) \overset{(5.29d)}{=} \begin{cases} \emptyset, & \text{if } \mathcal{A} \subseteq \delta(\mathcal{H}) \setminus \mathcal{J} \\ \mathcal{Y}^{(t)}, & \text{if } \mathcal{A} = \{j\}. \end{cases} \tag{5.61}$$

Moreover, for any $r \in \{1, \ldots, |\mathcal{J}| - 1\}$ and $t \in \{1, \ldots, i^*_{i,r} - 1\}$,

$$\delta(\mathcal{Y}^{(i,r;t)}) \overset{(5.29d)}{\subseteq} \{j^{(i,r)}\} \overset{\text{line 2}}{\subset} \mathcal{J}^{(i,r)} \overset{\text{Def. 5.4.1}}{\subset} \mathcal{J} \setminus \{j^{(i,0)}\} \overset{(5.35)}{=} \mathcal{J} \setminus \{j\}, \tag{5.62}$$

which implies $\gamma_{\mathcal{A}}(\mathcal{Y}^{(i,r;t)}) = \emptyset$ for both cases of $\mathcal{A}$ specified in (5.59).

Combining (5.60) with (5.61) and (5.62) yields the desired result (5.59). □

Lemma 5.5.3. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. The following relations hold:*

$$\delta(\mathcal{H} \setminus \mathcal{U}) \setminus \mathcal{J} \subseteq \bigcap_{i=1}^{i^*} \delta(\mathcal{H} \setminus \mathcal{V}_i), \tag{5.63a}$$

$$\mathcal{Y}^{(i)} \subset \bigcap_{t=1}^{i} \gamma_j(\mathcal{H} \setminus \mathcal{V}_t), \qquad i \in \{1, \ldots, i^*\}, \tag{5.63b}$$

$$j \in \bigcap_{i=1}^{i^*-1} \delta(\mathcal{H} \setminus \mathcal{V}_i), \qquad i^* \geq 2, \tag{5.63c}$$

$$\gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*}) = \emptyset \quad \Longrightarrow \quad i^* \geq 2, \tag{5.64a}$$

$$\gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*}) = \emptyset \quad \Longrightarrow \quad \mathcal{Y}^{(i^*-1)} = \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*-1}). \tag{5.64b}$$

Proof.

<u>(5.63a):</u> For each $i \in \{1, \ldots, i^*\}$, we observe that:

$$\begin{aligned} \{d \in \delta(\mathcal{H}) \colon \mathcal{V}_i \supseteq \gamma_d(\mathcal{H})\} &\subseteq \{d \in \delta(\mathcal{H}) \setminus \mathcal{J} \colon \mathcal{V}_i \supseteq \gamma_d(\mathcal{H})\} \cup \mathcal{J} \\ &= \{d \in \delta(\mathcal{H}) \setminus \mathcal{J} \colon \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{V}_i) \supseteq \gamma_d(\mathcal{H})\} \cup \mathcal{J} \\ &= \{d \in \delta(\mathcal{H}) \colon \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{V}_i) \supseteq \gamma_d(\mathcal{H})\} \cup \mathcal{J} \\ &\overset{(5.59)}{=} \{d \in \delta(\mathcal{H}) \colon \mathcal{U} \supseteq \gamma_d(\mathcal{H})\} \cup \mathcal{J}. \end{aligned} \tag{5.65}$$

Consequently, for each $i \in \{1, \ldots, i^*\}$, it follows that:

$$\begin{aligned}
\delta(\mathcal{H} \setminus \mathcal{V}_i) &\overset{(2.17f)}{=} \delta(\mathcal{H}) \setminus \{d \in \delta(\mathcal{H}) \colon \mathcal{V}_i \supseteq \gamma_d(\mathcal{H})\} \\
&\overset{(5.65)}{\supseteq} \delta(\mathcal{H}) \setminus \big(\{d \in \delta(\mathcal{H}) \colon \mathcal{U} \supseteq \gamma_d(\mathcal{H})\} \cup \mathcal{J}\big) \\
&\overset{(2.17f)}{=} \delta(\mathcal{H} \setminus \mathcal{U}) \setminus \mathcal{J}. \qquad (5.66)
\end{aligned}$$

(5.63b): For each $i \in \{1, \ldots, i^*\}$,

$$\begin{aligned}
\mathcal{Y}^{(i)} &\overset{\text{line 10}}{\subseteq} \mathcal{K}^{(i-1)} \overset{(5.31)}{=} \gamma_j(\mathcal{H}) \setminus \bigcup_{t=1}^{i-1} \mathcal{Y}^{(t)} = \gamma_j(\mathcal{H}) \setminus \bigcup_{k=1}^{i} \bigcup_{t=1}^{k-1} \mathcal{Y}^{(t)} \\
&\overset{(5.59)}{=} \gamma_j(\mathcal{H}) \setminus \bigcup_{k=1}^{i} \gamma_j(\mathcal{V}_k) \overset{\substack{(2.17a)\\(2.17c)}}{=} \gamma_j(\mathcal{H} \setminus \bigcup_{k=1}^{i} \mathcal{V}_k) \\
&= \gamma_j(\mathcal{H} \cap \Big(\bigcup_{k=1}^{i} \mathcal{V}_k\Big)^{\complement}) = \gamma_j(\bigcap_{k=1}^{i} (\mathcal{H} \setminus \mathcal{V}_k)) \\
&\overset{(2.17b)}{=} \bigcap_{k=1}^{i} \gamma_j(\mathcal{H} \setminus \mathcal{V}_k). \qquad (5.67)
\end{aligned}$$

(5.63c): Assume $i^* \geq 2$. For any $i \in \{1, \ldots, i^* - 1\}$, we have

$$\gamma_j(\mathcal{H} \setminus \mathcal{V}_i) \overset{(5.63b)}{\supset} \mathcal{Y}^{(i)} \overset{(5.29a)}{\neq} \emptyset, \qquad (5.68)$$

which implies $j \in \delta(\mathcal{H} \setminus \mathcal{V}_i)$.

(5.64): By (2.17c), the condition $\gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*}) = \emptyset$ is equivalent to

$$\gamma_j(\mathcal{H}) = \gamma_j(\mathcal{V}_{i^*}). \qquad (5.69)$$

Then we have

$$\emptyset \neq \gamma_j(\mathcal{H}) \overset{(5.69)}{=} \gamma_j(\mathcal{V}_{i^*}) \overset{(5.59)}{=} \bigcup_{i=1}^{i^*-1} \mathcal{Y}^{(i)}. \qquad (5.70)$$

If $i^* = 1$, the union in (5.70) is empty by convention, which yields a contradiction. Hence, $i^* \geq 2$, establishing (5.64a).

Furthermore, for $i^* \geq 2$, it follows that:

$$\begin{aligned}
\gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*-1}) &\overset{(2.17c)}{=} \gamma_j(\mathcal{H}) \setminus \gamma_j(\mathcal{V}_{i^*-1}) \\
&\overset{\substack{(5.70)\\(5.59)}}{=} \bigcup_{i=1}^{i^*-1} \mathcal{Y}^{(i)} \setminus \bigcup_{i=1}^{i^*-2} \mathcal{Y}^{(i)} \overset{(5.29c)}{=} \mathcal{Y}^{(i^*-1)}, \qquad (5.71)
\end{aligned}$$

which proves (5.64b).

□

Chapter 6

Correctness of the RDCA Algorithm

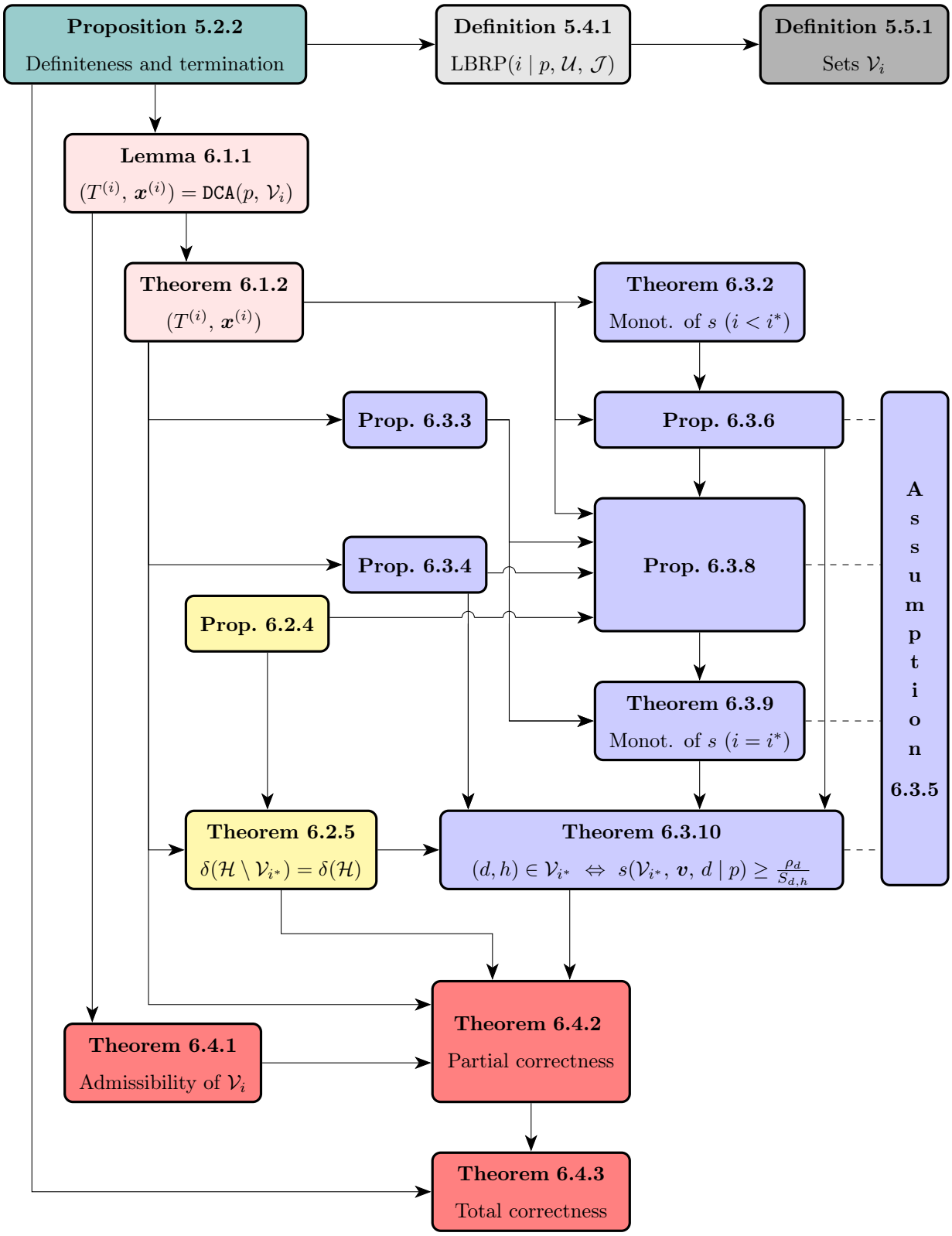
This chapter provides a formal proof that the `RDCA` algorithm solves the CPDA problem, thereby establishing its total correctness. As the most analytically intensive part of this thesis, the development builds directly upon the notational framework and structural properties established in Chapter 5.

Specifically, we establish the total correctness of $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ for the case $p \in \mathcal{P}$, $\mathcal{U} = \emptyset$, and $\mathcal{J} = \delta(\mathcal{H})$. Following the classic decomposition into partial correctness and termination (see Hoare [19]), our analysis verifies that the algorithm's output satisfies the optimality conditions (3.61)–(3.63) of Theorem 3.2.8. Each of the subsequent sections is dedicated to verifying a specific requirement from (3.61)–(3.63). These results culminate in Section 6.4, where we prove that all conditions of Theorem 3.2.8 are met, establishing partial correctness in Theorem 6.4.2. Finally, Theorem 6.4.3 asserts total correctness by encompassing both partial correctness and the termination results from the previous chapter.

All results from Chapter 5 remain in effect. In particular, the notational conventions for program variables (Sec. 5.1), the termination guarantees (Prop. 5.2.2), and the formal constructions of the Last-Branch Recursion Path (Def. 5.4.1) and the sets $\mathcal{V}_i$ (Def. 5.5.1) are pivotal to the optimality analysis developed here.

To assist the reader in navigating the complex interconnections between the various results presented in this chapter, Figure 6.0.1 provides a diagram summarizing the main theorems and their logical dependencies (including key foundational results from Chapter 5). To be explicit, the primary objective of this chapter is the proof of Theorem 6.4.3, which concludes the formal validation of the `RDCA` algorithm.

Figure 6.0.1: The main results of Chapter 6 and their logical interdependencies, including foundational definitions and results from Chapter 5. The background colours distinguish the various sections in which these results are established.



6.1. The Form of the Variable $(T, \boldsymbol{x})$

This section provides a characterization of the program variable $(T^{(i)}, \boldsymbol{x}^{(i)})$ in terms of the set $\mathcal{V}_i$ for $i \in \{1, \ldots, i^*\}$ (see Def. 5.5.1). The primary result of this characterization is given in Theorem 6.1.2.

Beyond this characterization, we derive several auxiliary properties of the $\boldsymbol{x}$ variable, most of which follow from the established characterization. While technical in nature, these results are essential for the developments in the subsequent sections.

Lemma 6.1.1. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. For any* $i \in \{1, \ldots, i^*\}$*, the model* p *and the set* $\mathcal{V}_i$ *defined by* (5.51) *together satisfy the input requirements of* `DCA`*. Moreover,*

$$(T^{(i)}, \boldsymbol{x}^{(i)}) = \texttt{DCA}(p, \mathcal{V}_i), \qquad i \in \{1, \ldots, i^*\}. \tag{6.1}$$

Proof. Fix an arbitrary $i \in \{1, \ldots, i^*\}$, and let $r := |\mathcal{J}| - 1$.

Consider the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ along the LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$. By Lemma 5.4.2, it follows that $|\mathcal{J}^{(i,r)}| = 1$. Consequently, the condition on line 5 is satisfied, and the assignment on line 6 yields:

$$(T^{(i,r;\tilde{i})}, \boldsymbol{x}^{(i,r;\tilde{i})}) = \texttt{DCA}(p, \mathcal{U}^{(i,r;\tilde{i}-1)}), \qquad \text{where } \tilde{i} := \begin{cases} i, & \text{if } r = 0 \\ i^*_{i,r}, & \text{if } r \geq 1. \end{cases} \tag{6.2}$$

By Proposition 5.2.2, p and $\mathcal{U}^{(i,r;\tilde{i}-1)}$ together satisfy the input requirements of `DCA`. Moreover,

$$\begin{aligned} (T^{(i,r;\tilde{i})}, \boldsymbol{x}^{(i,r;\tilde{i})}) &\overset{(5.46)}{=} (T^{(i)}, \boldsymbol{x}^{(i)}), \\ \mathcal{U}^{(i,r;\tilde{i}-1)} &\overset{(5.30b)}{=} \mathcal{U}^{(i,r)} \cup \bigcup_{t=1}^{\tilde{i}-1} \mathcal{Y}^{(i,r;t)} \overset{[1]}{=} \mathcal{V}_{i,r;\tilde{i}} \overset{(5.55)}{=} \mathcal{V}_i, \end{aligned} \tag{6.3}$$

where [1] follows from (5.53) for $r = |\mathcal{J}| - 1$, and the set $\mathcal{V}_{i,r;\tilde{i}}$ is as specified in Rem. 5.5.2.

Substituting (6.3) into (6.2) establishes the desired result (6.1). □

Lemma 6.1.1 is particularly noteworthy as it reveals a fundamental structural property of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ algorithm. It establishes that, regardless of the cardinality of the set $\mathcal{J}$, for every iteration i of the loop, the pair $(T^{(i)}, \boldsymbol{x}^{(i)})$ – the output of a potentially deeply nested recursion – is in fact computed by the base-case `DCA` algorithm applied to the set $\mathcal{V}_i$. This structural equivalence was already clearly visible in the numerical examples for `RDCA` in Section 4.2 of Chapter 4 (Figures 4.2.1–4.2.3). It was also suggested by the visual trace of the Last-Branch Recursion Path in Figures 5.4.1 and 5.4.2 in Chapter 5.

Following the definition of the DCA algorithm (specifically, line 8), Lemma 6.1.1 immediately implies that the allocations $x_{d,h}^{(i)}$ for strata $(d,h) \in \mathcal{V}_i$ are blocked at $N_{d,h}$. This property is formally stated in Theorem 6.1.2.

Theorem 6.1.2. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. Let the function* s *and the operator* Eigen *be defined by Definitions 3.2.3 and 3.2.2, respectively. For any* $i \in \{1, \ldots, i^*\}$*, it holds that*

$$T^{(i)} = \begin{cases} 0, & \text{if } \mathcal{V}_i = \mathcal{H} \\ \lambda_i, & \text{if } \mathcal{V}_i \subsetneq \mathcal{H}, \end{cases} \qquad x_{d,h}^{(i)} = \begin{cases} N_{d,h}, & (d,h) \in \mathcal{V}_i \\ s(\mathcal{V}_i, \boldsymbol{v}_i, d \mid p)\, A_{d,h}(p), & (d,h) \in \mathcal{H} \setminus \mathcal{V}_i, \end{cases} \tag{6.4}$$

where $\boldsymbol{x}^{(i)} = (x_{d,h}^{(i)}, (d,h) \in \mathcal{H})$*, and* $(\lambda_i, \boldsymbol{v}_i) := \mathrm{Eigen}(\mathcal{V}_i \mid p)$ *for* $\mathcal{V}_i \subsetneq \mathcal{H}$*.*

Proof. By Lemma 6.1.1, we have $(T^{(i)}, \boldsymbol{x}^{(i)}) = \texttt{DCA}(p, \mathcal{V}_i)$ for any $i \in \{1, \ldots, i^*\}$. The identities in (6.4) then follow directly from the definition of the DCA algorithm. □

Proposition 6.1.3. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. For any* $i \in \{1, \ldots, i^*\}$*, it holds that*

$$\sum_{h \in \eta_d(\mathcal{H})} \Big(\frac{1}{x_{d,h}^{(i)}} - \frac{1}{N_{d,h}} \Big) [A_{d,h}(p)]^2 = T^{(i)}, \qquad d \in \mathcal{D}_i := \begin{cases} \delta(\mathcal{H}), & \text{if } \mathcal{V}_i = \mathcal{H} \\ \delta(\mathcal{H} \setminus \mathcal{V}_i), & \text{if } \mathcal{V}_i \subsetneq \mathcal{H}, \end{cases} \tag{6.5}$$

where $\boldsymbol{x}^{(i)} = (x_{d,h}^{(i)}, (d,h) \in \mathcal{H})$*.*

Proof. By Lemma 6.1.1, $(T^{(i)}, \boldsymbol{x}^{(i)}) = \texttt{DCA}(p, \mathcal{V}_i)$ for any $i \in \{1, \ldots, i^*\}$. The identities in (6.5) then follow immediately from property (5.5d) of Lemma 5.2.1. □

Remark 6.1.1. In view of Lemma 6.1.1 and line 8 of the DCA algorithm, the set $\mathcal{V}_i$ represents the collection of strata $(d,h) \in \mathcal{H}$ that are *blocked* at their upper bounds $N_{d,h}$. Consequently, following Remark 2.3.1 with $\mathcal{B} = \mathcal{V}_i$, the set $\delta(\mathcal{H} \setminus \mathcal{V}_i)$ represents the domains that are not blocked by $\mathcal{V}_i$. When $\mathcal{V}_i \subsetneq \mathcal{H}$, Proposition 6.1.3 asserts that the equation in (6.5) is required to hold only for these unblocked domains; it is not necessarily satisfied for domains that are blocked by $\mathcal{V}_i$. This observation is analogous to Remark 5.2.1.

Proposition 6.1.4. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program and let* $i \in \{1, \ldots, i^*\}$*. The inclusion* $\mathcal{Y}^{(i)} \subseteq \gamma_j(\mathcal{H} \setminus \mathcal{V}_i)$ *holds, and for any* $(d,h) \in \gamma_j(\mathcal{H} \setminus \mathcal{V}_i)$*,*

$$(d,h) \in \mathcal{Y}^{(i)} \iff x_{d,h}^{(i)} \geq N_{d,h}, \tag{6.6a}$$

$$(d,h) \in \gamma_j(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)} \iff x_{d,h}^{(i)} < N_{d,h}, \tag{6.6b}$$

where $\boldsymbol{x}^{(i)} = (x_{d,h}^{(i)}, (d,h) \in \mathcal{H})$*.*

Proof. From line 10 of program,

$$\mathcal{Y}^{(i)} = \big\{(d,h) \in \mathcal{K}^{(i-1)} \colon\ x_{d,h}^{(i)} \geq N_{d,h}\big\}, \tag{6.7}$$

so that for all $(d,h) \in \mathcal{K}^{(i-1)}$,

$$\begin{aligned} (d,h) \in \mathcal{Y}^{(i)} \quad &\Longleftrightarrow \quad x_{d,h}^{(i)} \geq N_{d,h}, \\ (d,h) \in \mathcal{K}^{(i-1)} \setminus \mathcal{Y}^{(i)} \quad &\Longleftrightarrow \quad x_{d,h}^{(i)} < N_{d,h}. \end{aligned} \tag{6.8}$$

To conclude, observe that

$$\mathcal{K}^{(i-1)} \overset{(5.31)}{=} \gamma_j(\mathcal{H}) \setminus \bigcup_{t=1}^{i-1} \mathcal{Y}^{(t)} \overset{\substack{(5.59)\\(2.17c)}}{=} \gamma_j(\mathcal{H} \setminus \mathcal{V}_i). \tag{6.9}$$

□

Proposition 6.1.5. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. For any* $i \in \{1, \ldots, i^*\}$*, it holds that*

$$x_{d,h}^{(i)} > 0, \qquad (d,h) \in \mathcal{H}, \tag{6.10a}$$

$$x_{d,h}^{(i)} < N_{d,h}, \qquad (d,h) \in \gamma_{\mathcal{J}}(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)}, \tag{6.10b}$$

$$x_{d,h}^{(i)} \leq N_{d,h}, \qquad (d,h) \in \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}), \tag{6.10c}$$

$$x_{d,h}^{(i^*)} \leq N_{d,h}, \qquad (d,h) \in \gamma_{\mathcal{J}}(\mathcal{H}), \tag{6.10d}$$

where $\boldsymbol{x}^{(i)} = (x_{d,h}^{(i)},\ (d,h) \in \mathcal{H})$.

Proof. Let $i \in \{1, \ldots, i^*\}$.

<u>(6.10a):</u> The claim follows directly from (6.4) after referring to Remark 3.2.4.

<u>(6.10b):</u> First, observe that

$$\begin{aligned} \gamma_{\mathcal{J}}(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)} &\overset{\substack{(5.40)\\(2.17d)}}{=} \Big(\gamma_j(\mathcal{H} \setminus \mathcal{V}_i) \cup \bigcup_{r=1}^{|\mathcal{J}|-1} \gamma_{j^{(i,r)}}(\mathcal{H} \setminus \mathcal{V}_i)\Big) \setminus \mathcal{Y}^{(i)} \\ &\overset{[1]}{=} \Big(\gamma_j(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)}\Big) \cup \bigcup_{r=1}^{|\mathcal{J}|-1} \gamma_{j^{(i,r)}}(\mathcal{H} \setminus \mathcal{V}_i) \\ &\overset{(5.55)}{=} \Big(\gamma_j(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)}\Big) \cup \bigcup_{r=1}^{|\mathcal{J}|-1} \gamma_{j^{(i,r)}}(\mathcal{H} \setminus \mathcal{V}_{i,r;i_{i,r}^*}), \end{aligned} \tag{6.11}$$

where the set $\mathcal{V}_{i,r;i_{i,r}^*}$ is as described in Remark 5.5.2. The identity [1] is justified by the fact that for any $r \in \{1, \ldots, |\mathcal{J}|-1\}$ along the LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$,

$$\delta(\mathcal{Y}^{(i)}) \overset{(5.29d)}{\subseteq} \{j\} \overset{(5.35)}{=} \{j^{(i,0)}\} \overset{\text{Def. 5.4.1}}{\nsubseteq} \mathcal{J}^{(i,r)} \overset{\text{line 2}}{\ni} j^{(i,r)}, \tag{6.12}$$

which implies $\gamma_{j^{(i,r)}}(\mathcal{Y}^{(i)}) = \emptyset$.

Now, suppose $(d,h) \in \gamma_{\mathcal{J}}(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)}$. By (6.11), one of the following two cases must hold:

(i) $(d,h) \in \gamma_j(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)}$;

(ii) $|\mathcal{J}| \geq 2$ and there exists $r \in \{1, \ldots, |\mathcal{J}|-1\}$ such that

$$(d,h) \in \gamma_{j^{(i,r)}}(\mathcal{H} \setminus \mathcal{V}_{i,r;i^*_{i,r}}). \tag{6.13}$$

Case (i). This matches the premise of the forward implication in (6.6b) of Proposition 6.1.4, which immediately yields $x^{(i)}_{d,h} < N_{d,h}$, as wanted.

Case (ii). Fix r as in this case and consider the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ along the LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$. Relative to this invocation, the forward implication in (6.6b) for the final iteration $i^*_{i,r}$ is written as

$$(d,h) \in \gamma_{j^{(i,r)}}(\mathcal{H} \setminus \mathcal{V}_{i,r;i^*_{i,r}}) \setminus \mathcal{Y}^{(i,r;i^*_{i,r})} \quad \Longrightarrow \quad x^{(i,r;i^*_{i,r})}_{d,h} < N_{d,h}. \tag{6.14}$$

Since $\mathcal{Y}^{(i,r;i^*_{i,r})} \overset{(5.29b)}{=} \emptyset$, it follows from (6.13) that the premise of (6.14) is satisfied. Consequently, recalling that $\boldsymbol{x}^{(i,r;i^*_{i,r})} \overset{(5.46)}{=} \boldsymbol{x}^{(i)}$, we conclude that $x^{(i)}_{d,h} < N_{d,h}$, as required.

(6.10c): We have:

$$\begin{aligned} x^{(i)}_{d,h} &\overset{(6.10b)}{<} N_{d,h}, \qquad (d,h) \in \gamma_{\mathcal{J}}(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)}, \\ x^{(i)}_{d,h} &\overset{(6.4)}{=} N_{d,h}, \qquad (d,h) \in \mathcal{V}_i. \end{aligned} \tag{6.15}$$

Moreover,

$$\gamma_{\mathcal{J}}(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)} \supset \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)} \overset{\substack{(2.17c)\\(5.29d)}}{=} \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}) \setminus \mathcal{V}_i. \tag{6.16}$$

Combining (6.15) with (6.16) establishes the desired result (6.10c).

(6.10d): Using the relations in (6.15) for $i = i^*$, we obtain

$$x^{(i^*)}_{d,h} \leq N_{d,h}, \qquad (d,h) \in \big(\gamma_{\mathcal{J}}(\mathcal{H} \setminus \mathcal{V}_{i^*}) \setminus \mathcal{Y}^{(i^*)}\big) \cup \mathcal{V}_{i^*}. \tag{6.17}$$

Then, (6.10d) follows immediately by noting that

$$\gamma_{\mathcal{J}}(\mathcal{H} \setminus \mathcal{V}_{i^*}) \setminus \mathcal{Y}^{(i^*)} \overset{\substack{(5.29b)\\(2.17c)}}{=} \gamma_{\mathcal{J}}(\mathcal{H}) \setminus \mathcal{V}_{i^*}. \tag{6.18}$$

□

Proposition 6.1.6. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. If* $i^* \geq 2$ *then*

$$\sum_{(d,h)\in\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H})\setminus\mathcal{U}} \left(x_{d,h}^{(i^*-1)} - x_{d,h}^{(i^*)}\right) + \sum_{(d,h)\in\gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H})} \left(x_{d,h}^{(i^*-1)} - x_{d,h}^{(i^*)}\right)$$
$$+ \sum_{(d,h)\in\mathcal{Y}^{(i^*-1)}} \left(x_{d,h}^{(i^*-1)} - N_{d,h}\right) + \sum_{(d,h)\in\gamma_j(\mathcal{H}\setminus\mathcal{V}_{i^*})} \left(x_{d,h}^{(i^*-1)} - x_{d,h}^{(i^*)}\right) = 0. \quad (6.19)$$

Proof. As an immediate consequence of lines 6 and 8, in conjunction with (5.5c) and (5.16c) respectively, it follows that:

$$\sum_{(d,h)\in\mathcal{H}} x_{d,h}^{(i)} = n, \qquad i \in \{1, \ldots, i^*\}. \quad (6.20)$$

Assume henceforth that $i^* \geq 2$.

The set $\mathcal{H}$ admits the following decomposition:

$$\begin{aligned}\mathcal{H} &\overset{\substack{(2.17a)\\(2.17d)}}{=} \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}) \cup \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}) \cup \gamma_j(\mathcal{V}_{i^*}) \cup \gamma_j(\mathcal{H}\setminus\mathcal{V}_{i^*}) \\ &\overset{[1]}{=} \left(\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}) \setminus \mathcal{U}\right) \cup \mathcal{U} \cup \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}) \cup \gamma_j(\mathcal{V}_{i^*-1}) \cup \mathcal{Y}^{(i^*-1)} \cup \gamma_j(\mathcal{H}\setminus\mathcal{V}_{i^*}), \quad (6.21)\end{aligned}$$

where [1] follows from:

$$\mathcal{U} \overset{(5.30a)}{\subseteq} \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}), \quad (6.22a)$$
$$\gamma_j(\mathcal{V}_{i^*}) \overset{(5.59)}{=} \gamma_j(\mathcal{V}_{i^*-1}) \cup \mathcal{Y}^{(i^*-1)}. \quad (6.22b)$$

Note that the six subsets in the decomposition (6.21) are mutually disjoint; in particular,

$$\gamma_j(\mathcal{V}_{i^*-1}) \cap \mathcal{Y}^{(i^*-1)} \overset{\substack{(5.59)\\(5.29c)}}{=} \emptyset. \quad (6.23)$$

Moreover, by (6.4),

$$\begin{aligned} x_{d,h}^{(i^*-1)} &= x_{d,h}^{(i^*)}, & (d,h) &\in \mathcal{V}_{i^*-1} \cap \mathcal{V}_{i^*} \overset{(5.51)}{\supset} \mathcal{U}, \\ x_{d,h}^{(i^*-1)} &= x_{d,h}^{(i^*)}, & (d,h) &\in \gamma_j(\mathcal{V}_{i^*-1}) \overset{(6.22b)}{\subset} \gamma_j(\mathcal{V}_{i^*}), \\ x_{d,h}^{(i^*)} &= N_{d,h}, & (d,h) &\in \gamma_j(\mathcal{V}_{i^*}) \overset{(6.22b)}{\supset} \mathcal{Y}^{(i^*-1)}. \end{aligned} \quad (6.24)$$

Consider equation in (6.20) for $i = i^* - 1$ and $i = i^*$. Subtracting the latter from the former, component-wise, and applying the decomposition (6.21) along with the equalities in (6.24), yields the desired equality (6.19). □

Proposition 6.1.7. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. If* $i^* \geq 2$ *then for all* $d \in \delta(\mathcal{H} \setminus \mathcal{U}) \setminus \mathcal{J}$*, the following implications hold:*

$$\Big(\exists h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) : \; x_{d,h}^{(i^*-1)} < x_{d,h}^{(i^*)}\Big) \quad \Longrightarrow \quad \Big(\forall h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) \quad x_{d,h}^{(i^*-1)} < x_{d,h}^{(i^*)}\Big), \tag{6.25a}$$

$$\Big(\exists h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) : \; x_{d,h}^{(i^*-1)} \leq x_{d,h}^{(i^*)}\Big) \quad \Longrightarrow \quad \Big(\forall h \in \eta_d(\mathcal{H}) \quad x_{d,h}^{(i^*-1)} \leq x_{d,h}^{(i^*)}\Big). \tag{6.25b}$$

Proof. Suppose $i^* \geq 2$.

(6.25a): Assume that there exists $(d, h) \in \mathcal{H}$ such that

$$d \in \delta(\mathcal{H} \setminus \mathcal{U}) \setminus \mathcal{J} \quad \wedge \quad h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) \quad \wedge \quad x_{d,h}^{(i^*-1)} < x_{d,h}^{(i^*)}. \tag{6.26}$$

Then, by Corollary 2.3.2,

$$(d, h) \in \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}) \setminus \mathcal{U}. \tag{6.27}$$

This set lies in the region corresponding to one of the cases in Theorem 6.1.2:

$$\begin{aligned} \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}) \setminus \mathcal{U} &\overset{(5.59)}{=} \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}) \setminus \big(\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{V}_{i^*-1}) \cup \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{V}_{i^*})\big) \\ &\overset{\substack{(2.17a)\\(2.17c)}}{=} \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H} \setminus (\mathcal{V}_{i^*-1} \cup \mathcal{V}_{i^*})) \\ &\subset \mathcal{H} \setminus (\mathcal{V}_{i^*-1} \cup \mathcal{V}_{i^*}) = (\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap (\mathcal{H} \setminus \mathcal{V}_{i^*}). \end{aligned} \tag{6.28}$$

Therefore, applying (6.4) for $i \in \{i^* - 1, i^*\}$ and using (6.28), the inequality in (6.26) implies

$$s(\mathcal{V}_{i^*-1}, \boldsymbol{v}_{i^*-1}, d \mid p) < s(\mathcal{V}_{i^*}, \boldsymbol{v}_{i^*}, d \mid p), \tag{6.29}$$

where $(\lambda_i, \boldsymbol{v}_i) := \text{Eigen}(\mathcal{V}_i \mid p)$ for $i \in \{i^* - 1, i^*\}$.
Multiplying both sides of (6.29) by $A_{d,w}(p)$ for all $w \in \eta_d(\mathcal{H} \setminus \mathcal{U})$, we obtain:

$$s(\mathcal{V}_{i^*-1}, \boldsymbol{v}_{i^*-1}, d \mid p)\, A_{d,w}(p) < s(\mathcal{V}_{i^*}, \boldsymbol{v}_{i^*}, d \mid p)\, A_{d,w}(p), \quad w \in \eta_d(\mathcal{H}\setminus\mathcal{U}). \tag{6.30}$$

Notice the following inclusion relations:

$$\gamma_d(\mathcal{H} \setminus \mathcal{U}) \overset{[1]}{\subset} \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H} \setminus \mathcal{U}) \overset{\substack{(2.17c)\\(6.28)}}{\subset} (\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap (\mathcal{H} \setminus \mathcal{V}_{i^*}), \tag{6.31}$$

where inclusion [1] holds because $d \in \delta(\mathcal{H} \setminus \mathcal{U}) \setminus \mathcal{J} \subset \delta(\mathcal{H}) \setminus \mathcal{J}$ (recall (6.26)).
Applying (6.4) once more for $i \in \{i^* - 1, i^*\}$ and using (6.31), the inequalities in (6.30) can be written directly in terms of $x_{d,w}^{(i)}$, yielding

$$x_{d,w}^{(i^*-1)} < x_{d,w}^{(i^*)}, \qquad w \in \eta_d(\mathcal{H} \setminus \mathcal{U}). \tag{6.32}$$

(6.25b): Consider the following implication, to be established for all $d \in \delta(\mathcal{H} \setminus \mathcal{U}) \setminus \mathcal{J}$:

$$\left(\exists h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) : \; x_{d,h}^{(i^*-1)} \leq x_{d,h}^{(i^*)}\right) \quad \Longrightarrow \quad \left(\forall h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) \quad x_{d,h}^{(i^*-1)} \leq x_{d,h}^{(i^*)}\right). \tag{6.33}$$

The proof of (6.33) follows the same steps as that of (6.25a), replacing $<$ by $\leq$ throughout in (6.26), (6.29), (6.30), and (6.32). No other modifications are required.

To derive implication (6.25b) from (6.33), observe that by (6.4) for $i \in \{i^*-1, i^*\}$, we have:

$$x_{d,h}^{(i^*-1)} = x_{d,h}^{(i^*)}, \qquad (d,h) \in \mathcal{V}_{i^*-1} \cap \mathcal{V}_{i^*} \overset{(5.51)}{\supset} \mathcal{U}. \tag{6.34}$$

Thus, excluding $\mathcal{U}$ from the conclusion of implication in (6.33) does not affect its validity.

□

6.2. Domain Blockage

Definition 5.5.1 introduces the set $\mathcal{V}_{i^*} \subseteq \mathcal{H}$ within the context of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program. In this section, we demonstrate that condition (3.62b) from Theorem 3.2.8 is satisfied for $\mathcal{U}^* = \mathcal{V}_{i^*}$, provided that $\mathcal{J} = \delta(\mathcal{H})$ and $\mathcal{V}_{i^*} \subsetneq \mathcal{H}$. This result is formally established in Theorem 6.2.5.

Following the interpretation in Remark 6.1.1, set $\mathcal{V}_{i^*}$ can be viewed as the collection of strata $(d,h) \in \mathcal{H}$ that are *blocked* at their upper bounds $N_{d,h}$. Accordingly, $\delta(\mathcal{H} \setminus \mathcal{V}_{i^*})$ represents the domains that remain *unblocked* with respect to $\mathcal{V}_{i^*}$. Thus, in view of Remark 3.2.5, condition (3.62b) with $\mathcal{U}^* = \mathcal{V}_{i^*}$ asserts that, if $\mathcal{V}_{i^*} \subsetneq \mathcal{H}$, then no domain $d \in \delta(\mathcal{H})$ is blocked by $\mathcal{V}_{i^*}$ at the termination of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program. This interpretation motivates the title of this section.

We note that, purely out of curiosity, at this stage of the analysis we do not know if $x_{d,h}^{(i^*)} < N_{d,h}$ for all $(d,h) \in \mathcal{H} \setminus \mathcal{V}_{i^*}$; whether it holds or not is irrelevant for the present discussion (Theorem 6.3.10 will later show that it does hold when $\mathcal{J} = \delta(\mathcal{H})$).

Proposition 6.2.1. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. For any* $i \in \{1, \ldots, i^*\}$*, the following implication holds*

$$T^{(i)} \leq 0 \qquad \Longrightarrow \qquad \Big(\forall (d,h) \in \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}) \quad x^{(i)}_{d,h} = N_{d,h}\Big). \tag{6.35}$$

Proof. If $\mathcal{J} \setminus \{j\} = \emptyset$, then the implication (6.35) holds vacuously for any $i \in \{1, \ldots, i^*\}$.

Assume instead that $\mathcal{J} \setminus \{j\} \neq \emptyset$. The proof proceeds by contradiction. Suppose, toward a contradiction, that there exist $i \in \{1, \ldots, i^*\}$ and $(d,h) \in \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H})$ such that:

$$T^{(i)} \quad \leq \quad 0, \tag{6.36a}$$

$$x^{(i)}_{d,h} \quad \overset{\substack{(6.10a)\\(6.10c)}}{\in} \quad (0,\, N_{d,h}). \tag{6.36b}$$

Let $\mathcal{V}_i$ be as defined in Definition 5.5.1, and let

$$\mathcal{D}_i := \begin{cases} \delta(\mathcal{H}), & \text{if } \ \mathcal{V}_i = \mathcal{H} \\ \delta(\mathcal{H} \setminus \mathcal{V}_i), & \text{if } \ \mathcal{V}_i \subsetneq \mathcal{H}. \end{cases} \tag{6.37}$$

Only two cases are possible for the pair (d,h): either $d \in \mathcal{D}_i$ or $d \in \delta(\mathcal{H}) \setminus \mathcal{D}_i$. We consider them separately.

<u>$d \in \mathcal{D}_i$</u>

In this case, (6.5) implies

$$T^{(i)} = \sum_{w \in \eta_d(\mathcal{H})\setminus\{h\}} \Big(\frac{1}{x^{(i)}_{d,w}} - \frac{1}{N_{d,w}}\Big)[A_{d,w}(p)]^2 + \Big(\frac{1}{x^{(i)}_{d,h}} - \frac{1}{N_{d,h}}\Big)[A_{d,h}(p)]^2 > 0, \tag{6.38}$$

where the inequality follows from $x^{(i)}_{d,w} \in (0,\, N_{d,w}]$ for $w \in \eta_d(\mathcal{H}) \setminus \{h\}$ (see (6.10a), (6.10c)), and from (6.36b). Clearly, (6.38) contradicts (6.36a).

<u>$d \in \delta(\mathcal{H}) \setminus \mathcal{D}_i$</u>

In this case,

$$\emptyset \neq \delta(\mathcal{H}) \setminus \mathcal{D}_i = \delta(\mathcal{H}) \setminus \delta(\mathcal{H} \setminus \mathcal{V}_i) \overset{(2.17f)}{=} \{\widetilde{d} \in \delta(\mathcal{H}) \colon \ \mathcal{V}_i \supseteq \gamma_{\widetilde{d}}(\mathcal{H})\}. \tag{6.39}$$

Hence, $\mathcal{V}_i \supseteq \gamma_d(\mathcal{H})$, which implies $(d,h) \in \mathcal{V}_i$. Consequently, by (6.4), $x^{(i)}_{d,h} = N_{d,h}$, contradicting (6.36b).

In either case, we reach a contradiction with (6.36). Hence, the proposition is proven.

□

Proposition 6.2.2. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. The following implication holds*

$$\{j\} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) = \emptyset \quad \Longrightarrow \quad T^{(i^*)} \leq 0. \tag{6.40}$$

Proof. Assume the premise of (6.40) holds. Expressed via the function γ_j, this is equivalent to

$$\gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*}) = \emptyset. \tag{6.41}$$

By (5.64a), this implies

$$i^* \geq 2. \tag{6.42}$$

Furthermore, from (6.6a) and (6.4), it follows that

$$x_{d,h}^{(i^*-1)} \geq N_{d,h}, \qquad (d,h) \in \mathcal{Y}^{(i^*-1)} \cup \gamma_j(\mathcal{V}_{i^*-1}) \overset{[1]}{=} \gamma_j(\mathcal{H}), \tag{6.43}$$

where the equality [1] is a consequence of (5.64b) and (6.41), utilizing (2.17a).

Since $j \overset{(5.63c)}{\in} \delta(\mathcal{H} \setminus \mathcal{V}_{i^*-1})$, the following identity holds by Proposition 6.1.3:

$$T^{(i^*-1)} = \sum_{h \in \eta_j(\mathcal{H})} \Big(\frac{1}{x_{j,h}^{(i^*-1)}} - \frac{1}{N_{j,h}} \Big) [A_{j,h}(p)]^2 \overset{(6.43)}{\leq} 0. \tag{6.44}$$

Moreover, by Proposition 6.1.6, in view of (6.42) and (6.41), it follows that

$$n_1 + n_2 + n_3 = 0, \tag{6.45a}$$

where

$$\begin{aligned} n_1 &:= \sum_{(d,h) \in \gamma_{\delta(\mathcal{H}) \setminus \mathcal{J}}(\mathcal{H}) \setminus \mathcal{U}} \Big(x_{d,h}^{(i^*-1)} - x_{d,h}^{(i^*)} \Big), \\ n_2 &:= \sum_{(d,h) \in \gamma_{\mathcal{J} \setminus \{j\}}(\mathcal{H})} \Big(x_{d,h}^{(i^*-1)} - x_{d,h}^{(i^*)} \Big) \overset{\substack{(6.35)\\(6.44)}}{=} \sum_{(d,h) \in \gamma_{\mathcal{J} \setminus \{j\}}(\mathcal{H})} \Big(N_{d,h} - x_{d,h}^{(i^*)} \Big), \\ n_3 &:= \sum_{(d,h) \in \mathcal{Y}^{(i^*-1)}} \Big(x_{d,h}^{(i^*-1)} - N_{d,h} \Big). \end{aligned} \tag{6.45b}$$

We remark that the terms in the sum (6.45a) satisfy

$$n_2 \overset{(6.10c)}{\geq} 0 \qquad \text{and} \qquad n_3 \overset{(6.6a)}{\geq} 0, \tag{6.46}$$

which, in view of the equality in (6.45a), implies that $n_1 \leq 0$.

In summary, the premise of (6.40) yields (6.42), (6.44), and (6.45a). To prove the proposition, we analyze the inequality $n_1 \leq 0$ under two exhaustive cases: $\gamma_{\delta(\mathcal{H}) \setminus \mathcal{J}}(\mathcal{H}) \setminus \mathcal{U} \neq \emptyset$ and $\gamma_{\delta(\mathcal{H}) \setminus \mathcal{J}}(\mathcal{H}) \setminus \mathcal{U} = \emptyset$.

$\underline{\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H})\setminus\mathcal{U}\neq\emptyset}$. In this case, the inequality $n_1 \leq 0$ implies the existence of a pair $(d,h)\in\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H})\setminus\mathcal{U}$ such that

$$x_{d,h}^{(i^*-1)} \leq x_{d,h}^{(i^*)}. \tag{6.47}$$

Fix such a pair (d,h). By Corollary 2.3.2,

$$d\in\delta(\mathcal{H}\setminus\mathcal{U})\setminus\mathcal{J} \qquad \text{and} \qquad h\in\eta_d(\mathcal{H}\setminus\mathcal{U}). \tag{6.48}$$

Based on these properties, it follows from (5.63a) that

$$d\in\delta(\mathcal{H}\setminus\mathcal{V}_{i^*-1})\cap\delta(\mathcal{H}\setminus\mathcal{V}_{i^*}), \tag{6.49a}$$

while from (6.25b) one has

$$x_{d,w}^{(i^*-1)} \leq x_{d,w}^{(i^*)}, \qquad w\in\eta_d(\mathcal{H}). \tag{6.49b}$$

Altogether, observations (6.49) yield

$$\begin{aligned} 0 \overset{(6.44)}{\geq} T^{(i^*-1)} &\overset{\substack{(6.5)\\(6.49a)}}{=} \sum_{w\in\eta_d(\mathcal{H})}\Big(\frac{1}{x_{d,w}^{(i^*-1)}}-\frac{1}{N_{d,w}}\Big)[A_{d,w}(p)]^2 \\ &\overset{\substack{(6.49b)\\(6.10a)}}{\geq} \sum_{w\in\eta_d(\mathcal{H})}\Big(\frac{1}{x_{d,w}^{(i^*)}}-\frac{1}{N_{d,w}}\Big)[A_{d,w}(p)]^2 \overset{\substack{(6.49a)\\(6.5)}}{=} T^{(i^*)}. \end{aligned} \tag{6.50}$$

Hence, $T^{(i^*)}\leq 0$, which completes the proof of the proposition for this case.

$\underline{\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H})\setminus\mathcal{U}=\emptyset}$. In this case, $n_1=0$, and it follows from (6.45a) and (6.46) that $n_2=0$. This fact and (6.4) lead to the following identities, respectively:

$$\begin{aligned} x_{d,h}^{(i^*)} &\overset{\substack{n_2=0\\(6.10c)}}{=} N_{d,h}, \qquad (d,h)\in\gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}), \\ x_{d,h}^{(i^*)} &\overset{(6.4)}{=} N_{d,h}, \qquad (d,h)\in\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{V}_{i^*})\cup\gamma_j(\mathcal{V}_{i^*}). \end{aligned} \tag{6.51}$$

Moreover, we have:

$$\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{V}_{i^*}) \overset{(5.59)}{=} \mathcal{U} \overset{[1]}{=} \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}) \qquad \text{and} \qquad \gamma_j(\mathcal{V}_{i^*}) \overset{(6.41)}{=} \gamma_j(\mathcal{H}), \tag{6.52}$$

where [1] is a consequence of (5.30a) and the assumption defining the present case, i.e.,

$$\begin{aligned} \mathcal{U} &\subset \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}), \\ \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H})\setminus\mathcal{U} &= \emptyset. \end{aligned} \tag{6.53}$$

Combining the identities in (6.51) and (6.52) yields

$$x_{d,h}^{(i^*)} = N_{d,h}, \qquad (d,h)\in\mathcal{H}. \tag{6.54}$$

In light of (6.54), Proposition 6.1.3 implies $T^{(i^*)}=0$.

□

Proposition 6.2.3. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. For any nonempty* $\mathcal{A} \subseteq \mathcal{J}$*, the following implication holds*

$$\mathcal{A} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) = \emptyset \quad \Longrightarrow \quad T^{(i^*)} \leq 0. \tag{6.55}$$

Proof. Let $\emptyset \neq \mathcal{A} \subseteq \mathcal{J}$. Within the context of the LBRP$(i^* \mid p, \mathcal{U}, \mathcal{J})$, by Lemma 5.4.1,

$$\mathcal{A} \subseteq \bigcup_{r=0}^{|\mathcal{J}|-1} \{j^{(i^*,r)}\}. \tag{6.56}$$

Define the index set

$$R := \left\{ r \in \{0, \ldots, |\mathcal{J}| - 1\} \colon \; j^{(i^*,r)} \in \mathcal{A} \right\}. \tag{6.57}$$

It follows that

$$\begin{aligned} \mathcal{A} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) &= \Big(\bigcup_{r \in R} \{j^{(i^*,r)}\} \Big) \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \\ &= \bigcup_{r \in R} \Big(\{j^{(i^*,r)}\} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \Big) \\ &\overset{(5.55)}{=} \bigcup_{r \in R} \Big(\{j^{(i^*,r)}\} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*,r;i^*_{i^*,r}}) \Big), \end{aligned} \tag{6.58}$$

where the set $\mathcal{V}_{i^*,r;i^*_{i^*,r}}$ is as described in Remark 5.5.2. We note that $\mathcal{V}_{i^*,0;i^*_{i^*,0}} = \mathcal{V}_{i^*}$.

Under the premise of (6.55), the union in (6.58) is empty, which implies

$$\{j^{(i^*,r)}\} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*,r;i^*_{i^*,r}}) = \emptyset, \qquad r \in R. \tag{6.59}$$

Since $\mathcal{A} \neq \emptyset$, the set R is likewise nonempty. Fix an arbitrary $r \in R$ and consider the invocation $\texttt{RDCA}(p, \mathcal{U}^{(i^*,r)}, \mathcal{J}^{(i^*,r)})$ along the LBRP$(i^* \mid p, \mathcal{U}, \mathcal{J})$. Relative to this invocation, the implication (6.40) from Proposition 6.2.2 is written as

$$\{j^{(i^*,r)}\} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*,r;i^*_{i^*,r}}) = \emptyset \quad \Longrightarrow \quad T^{(i^*,r;i^*_{i^*,r})} \leq 0. \tag{6.60}$$

By (6.59), the premise of (6.60) is satisfied. Consequently, observing that

$$T^{(i^*,r;i^*_{i^*,r})} \overset{(5.46)}{=} T^{(i^*)}, \tag{6.61}$$

we conclude that $T^{(i^*)} \leq 0$. □

Proposition 6.2.4. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. The following implication holds*

$$\mathcal{J} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \neq \emptyset \quad \Longrightarrow \quad \mathcal{J} \subseteq \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}). \tag{6.62}$$

Proof. The proof proceeds by contradiction. Suppose that (6.62) does not hold. Then there exists $d \in \delta(\mathcal{H})$ such that

$$d \in \mathcal{J} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}), \tag{6.63a}$$

$$\mathcal{J} \nsubseteq \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}). \tag{6.63b}$$

Define $\mathcal{A} := \mathcal{J} \setminus \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})$. By construction, $\mathcal{A} \subset \mathcal{J}$ and $\mathcal{A} \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) = \emptyset$. Moreover, $\mathcal{A} \neq \emptyset$ by (6.63b). Hence, Proposition 6.2.3 yields

$$T^{(i^*)} \leq 0. \tag{6.64}$$

Since $d \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})$ according to (6.63a), it follows from Proposition 6.1.3 that

$$\sum_{h \in \eta_d(\mathcal{H})} \Big(\frac{1}{x_{d,h}^{(i^*)}} - \frac{1}{N_{d,h}} \Big) [A_{d,h}(p)]^2 = T^{(i^*)}. \tag{6.65}$$

Additionally, the condition $d \in \mathcal{J}$ implies:

$$x_{d,h}^{(i^*)} \overset{\substack{(6.10a)\\(6.10d)}}{\in} (0,\, N_{d,h}], \qquad h \in \eta_d(\mathcal{H}), \tag{6.66}$$

$$x_{d,h}^{(i^*)} \overset{\substack{(6.10b)\\(5.29b)}}{<} N_{d,h}, \qquad h \in \eta_d(\mathcal{H} \setminus \mathcal{V}_{i^*}). \tag{6.67}$$

Combining (6.64), (6.65), and (6.66) yields the identity

$$x_{d,h}^{(i^*)} = N_{d,h}, \qquad h \in \eta_d(\mathcal{H}). \tag{6.68}$$

This identity directly contradicts (6.67), since

$$\emptyset \overset{(6.63a)}{\neq} \eta_d(\mathcal{H} \setminus \mathcal{V}_{i^*}) \subset \eta_d(\mathcal{H}). \tag{6.69}$$

Therefore, the initial assumption must be false, and the implication (6.62) holds. □

We are now ready to state and prove the main result of this section, Theorem 6.2.5.

Theorem 6.2.5. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. If* $\mathcal{J} = \delta(\mathcal{H})$ *and* $\mathcal{V}_{i^*} \subsetneq \mathcal{H}$*, then*

$$\delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) = \delta(\mathcal{H}). \tag{6.70}$$

Proof. Assume $\mathcal{J} = \delta(\mathcal{H})$. Given the trivial inclusion

$$\delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \subseteq \delta(\mathcal{H}), \tag{6.71}$$

by Proposition 6.2.4, the following implication holds:

$$\delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \neq \emptyset \quad \Longrightarrow \quad \delta(\mathcal{H}) \subseteq \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}). \tag{6.72}$$

If $\mathcal{V}_{i^*} \subsetneq \mathcal{H}$, the premise of (6.72) is satisfied. Combining the resulting inclusion with (6.71) yields the desired equality (6.70). □

6.3. Monotonicity of the Function s

The primary goal of this section is to demonstrate that condition (3.62d) of Theorem 3.2.8 holds for $\mathcal{U}^* = \mathcal{V}_{i^*} \subsetneq \mathcal{H}$, where $\mathcal{V}_{i^*}$ is given by Definition 5.5.1 for the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ with $\mathcal{J} = \delta(\mathcal{H})$. This is formally established in Theorem 6.3.10. The proof of this claim relies on several auxiliary results, principally Theorems 6.3.2 and 6.3.9 concerning the monotonicity of the function s (recall Def. 3.2.3).

We begin with Corollary 6.3.1, which ensures that s is well-defined for the specified arguments under certain conditions.

Corollary 6.3.1. *Consider* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *and let* $i \in \{1, \ldots, i^*\}$. *Let* $\mathcal{V}_i$ *be the set defined in Definition 5.5.1. The following assertions hold:*

1. *If* $\mathcal{V}_i \subsetneq \mathcal{H}$, *then the pair* $(\lambda_i, \boldsymbol{v}_i) := \mathrm{Eigen}(\mathcal{V}_i \mid p)$ *is well-defined. Moreover, for every* $d \in \delta(\mathcal{H} \setminus \mathcal{V}_i)$, *the value* $s(\mathcal{V}_i, \boldsymbol{v}_i, d \mid p)$ *is well-defined.*
2. *If* $i^* \geq 2$ *and* $i < i^*$, *then the pair* $(\lambda_i, \boldsymbol{v}_i) := \mathrm{Eigen}(\mathcal{V}_i \mid p)$ *and the value* $s(\mathcal{V}_i, \boldsymbol{v}_i, j \mid p)$ *are well-defined.*

Proof. Assertion 1 follows directly from Theorem 6.1.2.

To prove assertion 2, assume that $i^* \geq 2$. By (5.63c), we have

$$j \in \bigcap_{i=1}^{i^*-1} \delta(\mathcal{H} \setminus \mathcal{V}_i). \tag{6.73}$$

Hence, $\mathcal{V}_i \subsetneq \mathcal{H}$ and $j \in \delta(\mathcal{H} \setminus \mathcal{V}_i)$ for all $i \in \{1, \ldots, i^* - 1\}$. The claim then follows from assertion 1. □

Abuse of notation for the function s. Let $\mathcal{V}_i$, for $i \in \{1, \ldots, i^*\}$, be the sets provided by Definition 5.5.1 in the context of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program. Throughout this section, we frequently encounter expressions involving $s(\mathcal{V}_i, \boldsymbol{v}_i, d \mid p)$ for $\mathcal{V}_i \subsetneq \mathcal{H}$, $(\lambda_i, \boldsymbol{v}_i) := \mathrm{Eigen}(\mathcal{V}_i \mid p)$, and $d \in \delta(\mathcal{H} \setminus \mathcal{V}_i)$. By Assertion 1 of Corollary 6.3.1, for these specific instances, the vector $\boldsymbol{v}_i$ is well-defined and uniquely determined by $\mathcal{V}_i$.

In view of this observation, we simplify the notation by writing $s(\mathcal{V}_i, d \mid p)$ as a shorthand for $s(\mathcal{V}_i, \boldsymbol{v}_i, d \mid p)$ for the parameters specified above. This convention is used throughout Section 6.3, with the exception of Theorem 6.3.10, which is later referenced in contexts requiring the full notation.

Theorem 6.3.2 (Monotonicity of s for $i < i^*$)**.** *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. If* $i^* \geq 3$*, then*

$$s(\mathcal{V}_{i-1},\ j \mid p) < s(\mathcal{V}_i,\ j \mid p), \qquad i \in \{2, \ldots, i^* - 1\}. \tag{6.74}$$

Proof. Suppose that $i^* \geq 3$ and let $i \in \{2, \ldots, i^* - 1\}$.

By (6.6), we have:

$$\begin{aligned} x_{d,h}^{(i-1)} &\overset{(6.6b)}{<} N_{d,h}, \qquad (d,h) \in \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i-1}) \setminus \mathcal{Y}^{(i-1)}, \\ x_{d,h}^{(i)} &\overset{(6.6a)}{\geq} N_{d,h}, \qquad (d,h) \in \mathcal{Y}^{(i)} \overset{(5.63b)}{\subset} \gamma_j(\mathcal{H} \setminus \mathcal{V}_i). \end{aligned} \tag{6.75}$$

Substituting $x_{d,h}^{(i-1)}$ and $x_{d,h}^{(i)}$ as given in (6.4), we obtain

$$\begin{aligned} s(\mathcal{V}_{i-1},\ d \mid p)\, A_{d,h}(p) &< N_{d,h}, \qquad (d,h) \in \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i-1}) \setminus \mathcal{Y}^{(i-1)}, \\ s(\mathcal{V}_i,\ d \mid p)\, A_{d,h}(p) &\geq N_{d,h}, \qquad (d,h) \in \mathcal{Y}^{(i)}. \end{aligned} \tag{6.76}$$

Since

$$\mathcal{Y}^{(i)} \overset{\substack{(5.63b)\\(5.29c)}}{\subset} \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i-1}) \setminus \mathcal{Y}^{(i-1)}, \tag{6.77}$$

the inequalities in (6.76) imply

$$s(\mathcal{V}_{i-1},\ d \mid p) < s(\mathcal{V}_i,\ d \mid p), \qquad d \in \delta(\mathcal{Y}^{(i)}). \tag{6.78}$$

The desired result (6.74) now follows from the fact that $\delta(\mathcal{Y}^{(i)}) = \{j\}$ (see (5.29a), (5.29d)). □

To establish the desired monotonicity in Theorem 6.3.2, we relied on condition (5.29a), which ensures that $\mathcal{Y}^{(i)} \neq \emptyset$ for $i \in \{1, \ldots, i^* - 1\}$. The case $i = i^*$ cannot be handled by this argument because $\mathcal{Y}^{(i^*)} \overset{(5.29b)}{=} \emptyset$. Therefore, a different proof strategy is required. This case is resolved in Theorem 6.3.9, which extends the monotonicity result to include $i = i^*$. The proof of Theorem 6.3.9 is based on several auxiliary results – Propositions 6.3.3, 6.3.4, 6.3.6, and 6.3.8 – which are presented below.

Proposition 6.3.3. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. Suppose that:*

(i) $i^* \geq 2$,

(ii) $j \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})$,

(iii) $s(\mathcal{V}_{i^*-1},\ j \mid p) > s(\mathcal{V}_{i^*},\ j \mid p)$.

Then the following statements hold:

$$\exists d \in \delta(\mathcal{H} \setminus \mathcal{U}) \setminus \{j\} :\ \exists h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) :\ x_{d,h}^{(i^*-1)} < x_{d,h}^{(i^*)}, \tag{6.79a}$$

$$\forall d \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})\ \ \exists h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) :\ x_{d,h}^{(i^*-1)} > x_{d,h}^{(i^*)}. \tag{6.79b}$$

Proof. We first observe that, by Corollary 6.3.1 and conditions *(i)–(ii)*, the function values appearing in condition *(iii)* are well-defined.

From (6.4), we recall the following identity for $i \in \{i^* - 1,\, i^*\}$:

$$x_{d,h}^{(i)} = s(\mathcal{V}_i,\, d \mid p)\, A_{d,h}(p), \qquad (d,h) \in \gamma_j(\mathcal{H} \setminus \mathcal{V}_i). \tag{6.80a}$$

Moreover,

$$\emptyset \overset{(ii)}{\neq} \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*}) \overset{\substack{(2.17c)\\(5.59)}}{\subset} \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*-1}). \tag{6.80b}$$

These properties will be utilized later in the proof.

<u>(6.79a):</u> The following set equalities hold:

$$\begin{aligned}
\big(\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}) \setminus \mathcal{U}\big) \cup \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}) &= \big(\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}) \cup \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H})\big) \setminus \big(\mathcal{U} \setminus \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H})\big) \\
&\overset{\substack{(2.17d)\\(5.30a)}}{=} \gamma_{(\delta(\mathcal{H})\setminus\mathcal{J})\cup(\mathcal{J}\setminus\{j\})}(\mathcal{H}) \setminus \mathcal{U} \\
&= \gamma_{\delta(\mathcal{H})\setminus\{j\}}(\mathcal{H}) \setminus \mathcal{U}, && (6.81a)\\
\big(\gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{H}) \setminus \mathcal{U}\big) \cap \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}) &\overset{(2.17e)}{=} \gamma_{(\delta(\mathcal{H})\setminus\mathcal{J})\cap(\mathcal{J}\setminus\{j\})}(\mathcal{H}) \setminus \mathcal{U} = \emptyset. && (6.81b)
\end{aligned}$$

Applying Proposition 6.1.6, together with (6.81) and (6.80), we obtain:

$$\begin{aligned}
&\sum_{(d,h)\in\gamma_{\delta(\mathcal{H})\setminus\{j\}}(\mathcal{H})\setminus\mathcal{U}} \Big(x_{d,h}^{(i^*-1)} - x_{d,h}^{(i^*)}\Big) + \sum_{(d,h)\in\mathcal{Y}^{(i^*-1)}} \Big(x_{d,h}^{(i^*-1)} - N_{d,h}\Big) \\
&\qquad + \Big(s(\mathcal{V}_{i^*-1},\, j \mid p) - s(\mathcal{V}_{i^*},\, j \mid p)\Big) \sum_{(d,h)\in\gamma_j(\mathcal{H}\setminus\mathcal{V}_{i^*})} A_{d,h}(p) = 0. \qquad (6.82)
\end{aligned}$$

Due to (6.6a), assumption *(iii)*, and the fact that $\gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*}) \overset{(6.80b)}{\neq} \emptyset$, equality (6.82) implies:

$$\sum_{(d,h)\in\gamma_{\delta(\mathcal{H})\setminus\{j\}}(\mathcal{H})\setminus\mathcal{U}} \Big(x_{d,h}^{(i^*-1)} - x_{d,h}^{(i^*)}\Big) < 0. \tag{6.83}$$

Consequently, the following holds:

$$\exists (d,h) \in \gamma_{\delta(\mathcal{H})\setminus\{j\}}(\mathcal{H}) \setminus \mathcal{U} :\ x_{d,h}^{(i^*-1)} < x_{d,h}^{(i^*)}. \tag{6.84}$$

By Corollary 2.3.2, statement (6.84) is equivalently expressed as

$$\exists d \in \delta(\mathcal{H} \setminus \mathcal{U}) \setminus \{j\} :\ \exists h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) :\ x_{d,h}^{(i^*-1)} < x_{d,h}^{(i^*)}. \tag{6.85}$$

(6.79b): By (6.5) and the fact that $j \in \delta(\mathcal{H} \setminus \mathcal{V}_i)$ for $i \in \{i^* - 1,\, i^*\}$ (see (5.63c) and *(ii)*), we have:

$$T^{(i)} = \sum_{h \in \eta_j(\mathcal{H})} \frac{[A_{j,h}(p)]^2}{x_{j,h}^{(i)}} - c_j(p), \qquad i \in \{i^* - 1,\, i^*\}. \tag{6.86}$$

Observe that the set $\gamma_j(\mathcal{H})$ decomposes as:

$$\begin{aligned} \gamma_j(\mathcal{H}) &\overset{(2.17a)}{=} \gamma_j(\mathcal{V}_{i^*}) \cup \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*}) \\ &\overset{(5.59)}{=} \gamma_j(\mathcal{V}_{i^*-1}) \cup \mathcal{Y}^{(i^*-1)} \cup \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*}), \end{aligned} \tag{6.87}$$

where the three subsets in (6.87) are mutually disjoint; in particular,

$$\gamma_j(\mathcal{V}_{i^*-1}) \cap \mathcal{Y}^{(i^*-1)} \overset{\substack{(5.59)\\(5.29c)}}{=} \emptyset. \tag{6.88}$$

Moreover, (6.4) gives:

$$\begin{aligned} x_{d,h}^{(i^*-1)} &= x_{d,h}^{(i^*)}, \qquad (d,h) \in \gamma_j(\mathcal{V}_{i^*-1} \cap \mathcal{V}_{i^*}) \overset{\substack{(2.17b)\\(5.59)}}{=} \gamma_j(\mathcal{V}_{i^*-1}), \\ x_{d,h}^{(i^*)} &= N_{d,h}, \qquad (d,h) \in \mathcal{V}_{i^*} \overset{(5.51)}{\supset} \mathcal{Y}^{(i^*-1)}. \end{aligned} \tag{6.89}$$

By subtracting (6.86) for $i = i^*$ from the case $i = i^* - 1$ and applying the established properties, we obtain

$$\begin{aligned} T^{(i^*-1)} - T^{(i^*)} &\overset{\substack{(6.86)\\(6.87)}}{=} \sum_{(d,h) \in \gamma_j(\mathcal{V}_{i^*-1})} \Big(\frac{1}{x_{d,h}^{(i^*-1)}} - \frac{1}{x_{d,h}^{(i^*)}}\Big)[A_{d,h}(p)]^2 \\ &\quad + \sum_{(d,h) \in \mathcal{Y}^{(i^*-1)}} \Big(\frac{1}{x_{d,h}^{(i^*-1)}} - \frac{1}{x_{d,h}^{(i^*)}}\Big)[A_{d,h}(p)]^2 \\ &\quad + \sum_{(d,h) \in \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*})} \Big(\frac{1}{x_{d,h}^{(i^*-1)}} - \frac{1}{x_{d,h}^{(i^*)}}\Big)[A_{d,h}(p)]^2 \\ &\overset{\substack{(6.89)\\(6.80)}}{=} \sum_{(d,h) \in \mathcal{Y}^{(i^*-1)}} \Big(\frac{1}{x_{d,h}^{(i^*-1)}} - \frac{1}{N_{d,h}}\Big)[A_{d,h}(p)]^2 \\ &\quad + \Big(\frac{1}{s(\mathcal{V}_{i^*-1},\, j \mid p)} - \frac{1}{s(\mathcal{V}_{i^*},\, j \mid p)}\Big) \sum_{(d,h) \in \gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*})} A_{d,h}(p). \end{aligned} \tag{6.90}$$

Furthermore, from (6.6a), it follows that

$$\sum_{(d,h) \in \mathcal{Y}^{(i^*-1)}} \Big(\frac{1}{x_{d,h}^{(i^*-1)}} - \frac{1}{N_{d,h}}\Big)[A_{d,h}(p)]^2 \le 0. \tag{6.91}$$

Combining (6.91) with assumption *(iii)* and the fact that $\gamma_j(\mathcal{H} \setminus \mathcal{V}_{i^*}) \overset{(6.80b)}{\neq} \emptyset$, the equality in (6.90) yields

$$T^{(i^*-1)} - T^{(i^*)} < 0. \tag{6.92}$$

Invoking (6.5) once more for $i \in \{i^* - 1,\, i^*\}$ and using (6.92), we deduce,

$$\forall d \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \quad \sum_{h \in \eta_d(\mathcal{H})} \Big(\frac{1}{x_{d,h}^{(i^*-1)}} - \frac{1}{x_{d,h}^{(i^*)}} \Big) [A_{d,h}(p)]^2 < 0. \tag{6.93}$$

Since $x_{d,h}^{(i^*-1)},\, x_{d,h}^{(i^*)} > 0$ for $(d,h) \in \mathcal{H}$ (see (6.10a)), it follows from (6.93) that

$$\forall d \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \quad \exists h \in \eta_d(\mathcal{H}) : \; x_{d,h}^{(i^*-1)} > x_{d,h}^{(i^*)}. \tag{6.94}$$

Finally, observing that

$$x_{d,h}^{(i^*-1)} \overset{(6.4)}{=} x_{d,h}^{(i^*)}, \qquad (d,h) \in \mathcal{V}_{i^*-1} \cap \mathcal{V}_{i^*} \overset{(5.51)}{\supset} \mathcal{U}, \tag{6.95}$$

the assertion in (6.94) can be refined to

$$\forall d \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \quad \exists h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) : \; x_{d,h}^{(i^*-1)} > x_{d,h}^{(i^*)}. \tag{6.96}$$

□

Proposition 6.3.4. *Consider the* RDCA$(p, \mathcal{U}, \mathcal{J})$ *program. Let* $\mathcal{A} \subseteq \mathcal{J}$ *and define*

$$I := \begin{cases} \{i^*\}, & \text{if } \; j \in \mathcal{A} \\ \{1, \ldots, i^*\}, & \text{if } \; j \notin \mathcal{A}. \end{cases} \tag{6.97}$$

Then, for all $i \in I$ *and all* $(d,h) \in \gamma_{\delta(\mathcal{H} \setminus \mathcal{V}_i)}(\mathcal{H})$*, the following implication holds:*

$$(d,h) \in \gamma_{\delta(\mathcal{H} \setminus \mathcal{V}_i) \cap \mathcal{A}}(\mathcal{H}) \setminus \mathcal{V}_i \quad \Longrightarrow \quad s(\mathcal{V}_i,\, d \mid p) < \tfrac{\rho_d}{S_{d,h}}. \tag{6.98}$$

Proof. Let $i \in I$.

Observe that

$$\gamma_{\mathcal{A}}(\mathcal{Y}^{(i)}) = \emptyset. \tag{6.99}$$

Indeed, in the first case of (6.97), $\mathcal{Y}^{(i^*)} \overset{(5.29b)}{=} \emptyset$. In the second case, since $j \notin \mathcal{A}$, the result follows from $\delta(\mathcal{Y}^{(i)}) \overset{(5.29d)}{\subseteq} \{j\}$.

Applying (6.99), we obtain

$$\begin{aligned} \gamma_{\delta(\mathcal{H} \setminus \mathcal{V}_i) \cap \mathcal{A}}(\mathcal{H}) \setminus \mathcal{V}_i &\overset{(2.17c)}{=} \gamma_{\delta(\mathcal{H} \setminus \mathcal{V}_i) \cap \mathcal{A}}(\mathcal{H} \setminus \mathcal{V}_i) \\ &\overset{(6.99)}{=} \gamma_{\delta(\mathcal{H} \setminus \mathcal{V}_i) \cap \mathcal{A}}(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)} \overset{\mathcal{A} \subset \mathcal{J}}{\subset} \gamma_{\mathcal{J}}(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)}. \end{aligned} \tag{6.100}$$

Furthermore, by (6.10b) and (6.4), we have:

$$s(\mathcal{V}_i,\, d \mid p)\, A_{d,h}(p) < N_{d,h}, \qquad (d,h) \in \gamma_{\mathcal{J}}(\mathcal{H} \setminus \mathcal{V}_i) \setminus \mathcal{Y}^{(i)}. \tag{6.101}$$

Combining (6.100) with (6.101), and using the identity $\frac{N_{d,h}}{A_{d,h}(p)} = \frac{\rho_d}{S_{d,h}}$, we obtain the desired implication (6.98). □

The subsequent propositions rely on a specific monotonicity condition for the function s. To ensure a modular and compact presentation, this condition is formalized as Assumption 6.3.5, parameterized by $u \in \mathbb{N}$.

Assumption 6.3.5. Let $u \in \mathbb{N}$ be given. For the program $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ with any valid input parameters satisfying $|\mathcal{J}| \leq u$, the following implication holds:

$$\Big(i^* \geq 2 \quad \wedge \quad j \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})\Big) \quad \Longrightarrow \quad s(\mathcal{V}_{i^*-1},\, j \mid p) \leq s(\mathcal{V}_{i^*},\, j \mid p). \tag{6.102}$$

Regarding Assumption 6.3.5, Corollary 6.3.1 states that the premise of implication (6.102) guarantees that the values of $s(\mathcal{V}_i,\, j \mid p)$ are well-defined for $i \in \{i^* - 1,\, i^*\}$, thereby confirming the validity of this formulation.

Proposition 6.3.6. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. Let*

$$\mathcal{A} \in \begin{cases} \{\mathcal{J}\}, & \text{if } |\mathcal{J}| = 1 \\ \{\mathcal{J},\ \mathcal{J} \setminus \{j\}\}, & \text{if } |\mathcal{J}| \geq 2, \end{cases} \tag{6.103}$$

and define

$$I := \begin{cases} \{i^*\}, & \text{if } \mathcal{A} = \mathcal{J} \\ \{1, \ldots, i^*\}, & \text{if } \mathcal{A} = \mathcal{J} \setminus \{j\}. \end{cases} \tag{6.104}$$

If Assumption 6.3.5 holds with $u = |\mathcal{A}|$, then for all $i \in I$ and all $(d, h) \in \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_i)}(\mathcal{H})$, we have:

$$(d, h) \in \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_i)\cap\mathcal{A}}(\mathcal{H}) \cap \mathcal{V}_i \quad \Longrightarrow \quad s(\mathcal{V}_i,\, d \mid p) \geq \tfrac{\rho_d}{S_{d,h}}. \tag{6.105}$$

Proof. Fix an arbitrary $i \in I$.

If $\gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_i)\cap\mathcal{A}}(\mathcal{H}) \cap \mathcal{V}_i = \emptyset$, then the implication (6.105) holds vacuously.

Assume instead that this intersection is nonempty, and fix an arbitrary element

$$(d, h) \in \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_i)\cap\mathcal{A}}(\mathcal{H}) \cap \mathcal{V}_i. \tag{6.106}$$

To facilitate the analysis of this nontrivial case, we organize the remainder of the proof into three parts.

Part 1: Identification of the indices r and t.

By the definition (5.51) of $\mathcal{V}_i$ and the fact that $\gamma_{\mathcal{A}}(\mathcal{U}) \overset{(5.30a)}{=} \emptyset$, we have

$$\gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_i)\cap\mathcal{A}}(\mathcal{H}) \cap \mathcal{V}_i \subseteq \bigcup_{r=0}^{|\mathcal{J}|-1} \bigcup_{t=1}^{\tilde{i}_r - 1} \mathcal{Y}^{(i,r;t)}, \qquad \text{where} \quad \tilde{i}_r := \begin{cases} i, & \text{if } r = 0 \\ i^*_{i,r}, & \text{if } r \geq 1. \end{cases} \tag{6.107}$$

Consequently, there exist $r \in \{0, \ldots, |\mathcal{J}| - 1\}$ and $t \in \{1, \ldots, \tilde{i}_r - 1\}$ such that

$$(d, h) \in \mathcal{Y}^{(i,r;t)}. \tag{6.108}$$

Together with the inclusion $\delta(\mathcal{Y}^{(i,r;t)}) \overset{(5.29d)}{\subseteq} \{j^{(i,r)}\}$, this implies

$$d = j^{(i,r)}. \tag{6.109}$$

We note the following properties of the recursion level r:

$$\mathcal{A} = \mathcal{J} \setminus \{j\} \quad \Longrightarrow \quad r \geq 1, \tag{6.110a}$$

$$r = 0 \quad \Longrightarrow \quad i = i^*. \tag{6.110b}$$

The first implication follows because if $j \notin \mathcal{A}$, then (6.106) implies $d \neq j = j^{(i,0)}$; thus, $r \geq 1$ by (6.109). The second follows by contraposition of the first: if $r = 0$, then $\mathcal{A} = \mathcal{J}$, which by the definition of I requires $i = i^*$.

Implication (6.110b), combined with the third notational identity in (5.35), ensures that $\tilde{i}_r = i^*_{i,r}$. Therefore, $t \in \{1, \ldots, i^*_{i,r} - 1\}$, and consequently,

$$i^*_{i,r} \geq 2. \tag{6.111}$$

Part 2: Analysis of the $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ invocation.

For the $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ invocation along the LBRP$(i \mid p, \mathcal{U}, \mathcal{J})$, let the sets $\mathcal{V}_{i,r;g}$, $g \in \{1, \ldots, i^*_{i,r}\}$, be defined as explained in Remark 5.5.1. Then, by (5.63b), for $g = t$ we have

$$\mathcal{Y}^{(i,r;t)} \subset \mathcal{H} \setminus \mathcal{V}_{i,r;t}. \tag{6.112}$$

Consequently, since $(d, h) \in \mathcal{Y}^{(i,r;t)}$ (recall (6.108)), we have

$$\begin{aligned} x^{(i,r;t)}_{d,h} &\overset{(6.4)}{=} s(\mathcal{V}_{i,r;t},\, d \mid p)\, A_{d,h}(p), \\ x^{(i,r;t)}_{d,h} &\overset{(6.6a)}{\geq} N_{d,h}. \end{aligned} \tag{6.113}$$

Combining both parts of (6.113) gives

$$s(\mathcal{V}_{i,r;t},\, d \mid p) \geq \tfrac{N_{d,h}}{A_{d,h}(p)} = \tfrac{\rho_d}{S_{d,h}}. \tag{6.114}$$

For later reference, by (5.55) and using the fact that $\tilde{i}_r = i^*_{i,r}$, we have

$$\mathcal{V}_i = \mathcal{V}_{i,r;i^*_{i,r}}. \tag{6.115}$$

Part 3: Invocation of the monotonicity of s.

Observe that $|\mathcal{J}| - r \leq |\mathcal{A}|$. This is evident if $\mathcal{A} = \mathcal{J}$, and for $\mathcal{A} = \mathcal{J} \setminus \{j\}$ it follows from $r \overset{(6.110a)}{\geq} 1$. Consequently, by (5.43), we obtain:

$$|\mathcal{J}^{(i,r)}| \leq |\mathcal{A}|. \tag{6.116}$$

Let Assumption 6.3.5 hold for $u = |\mathcal{A}|$. Then, since $|\mathcal{J}^{(i,r)}| \leq u$ for this choice of u, the following implication holds for the $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ program:

$$\Big(i^*_{i,r} \geq 2 \ \wedge \ j^{(i,r)} \in \delta(\mathcal{H} \setminus \mathcal{V}_{i,r;i^*_{i,r}})\Big) \implies$$
$$s(\mathcal{V}_{i,r;i^*_{i,r}-1},\, j^{(i,r)} \mid p) \leq s(\mathcal{V}_{i,r;i^*_{i,r}},\, j^{(i,r)} \mid p). \tag{6.117}$$

The premise of (6.117) is satisfied because of (6.111) and

$$j^{(i,r)} \overset{(6.109)}{=} d \overset{(6.106)}{\in} \delta(\mathcal{H} \setminus \mathcal{V}_i) \overset{(6.115)}{=} \delta(\mathcal{H} \setminus \mathcal{V}_{i,r;i^*_{i,r}}). \tag{6.118}$$

Combining the conclusion of (6.117) with Theorem 6.3.2 applied to the $\texttt{RDCA}(p, \mathcal{U}^{(i,r)}, \mathcal{J}^{(i,r)})$ program, we get

$$s(\mathcal{V}_{i,r;g},\, j^{(i,r)} \mid p) \leq s(\mathcal{V}_{i,r;i^*_{i,r}},\, j^{(i,r)} \mid p), \qquad g \in \{1, \ldots, i^*_{i,r} - 1\}. \tag{6.119}$$

In particular, for $g = t$ identified in Part 1, and in view of (6.109), we obtain

$$s(\mathcal{V}_{i,r;t},\, d \mid p) \leq s(\mathcal{V}_{i,r;i^*_{i,r}},\, d \mid p). \tag{6.120}$$

The desired conclusion then follows from

$$\frac{\rho_d}{S_{d,h}} \overset{\substack{(6.114)\\(6.120)}}{\leq} s(\mathcal{V}_{i,r;i^*_{i,r}},\, d \mid p) \overset{(6.115)}{=} s(\mathcal{V}_i,\, d \mid p). \tag{6.121}$$

□

We now state Proposition 6.3.8, which shows that certain conditions on the variables of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program cannot all hold simultaneously. This result plays a key role in the proof of Theorem 6.3.9, which proceeds by contradiction. Before proceeding, Lemma 6.3.7 records an observation that will be used in the proof of Proposition 6.3.8.

Lemma 6.3.7. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program and let* $i \in \{i^* - 1, i^*\}$. *For any* $(d, h) \in \mathcal{V}_i$, *the following implication holds:*

$$\Big(d \neq j \quad \wedge \quad h \in \eta_d(\mathcal{H} \setminus \mathcal{U})\Big) \quad \implies \quad (d, h) \in \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}). \tag{6.122}$$

Proof. First, observe that

$$\gamma_{\delta(\mathcal{H})\setminus\{j\}}(\mathcal{V}_i) \overset{(2.17d)}{=} \gamma_{\delta(\mathcal{H})\setminus\mathcal{J}}(\mathcal{V}_i) \cup \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{V}_i) \overset{(5.59)}{=} \mathcal{U} \cup \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{V}_i). \tag{6.123}$$

Let $(d,h) \in \mathcal{V}_i$ with $d \neq j$. Then

$$(d,h) \in \gamma_{\delta(\mathcal{H})\setminus\{j\}}(\mathcal{V}_i) \overset{(6.123)}{=} \mathcal{U} \cup \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{V}_i). \tag{6.124}$$

Furthermore, if $h \in \eta_d(\mathcal{H} \setminus \mathcal{U})$, then clearly

$$(d,h) \in \gamma_d(\mathcal{H} \setminus \mathcal{U}) \overset{(2.17c)}{=} \gamma_d(\mathcal{H}) \setminus \mathcal{U}. \tag{6.125}$$

Combining (6.124) and (6.125), we obtain

$$(d,h) \in \big(\mathcal{U} \cup \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{V}_i)\big) \cap \big(\gamma_d(\mathcal{H}) \setminus \mathcal{U}\big) \subset \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{V}_i) \subset \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}). \tag{6.126}$$

□

Proposition 6.3.8. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. The following conditions cannot all hold simultaneously:*

(i) $i^* \geq 2$,

(ii) $j \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})$,

(iii) $s(\mathcal{V}_{i^*-1},\, j \mid p) > s(\mathcal{V}_{i^*},\, j \mid p)$,

(iv) $|\mathcal{J}| \geq 2$,

(v) Assumption 6.3.5 holds for $u = |\mathcal{J}| - 1$.

Proof. First, observe that by Corollary 6.3.1 and conditions *(i)–(ii)*, the function values in condition *(iii)* are well-defined.

The proof of the proposition proceeds by contradiction. Specifically, assuming that conditions *(i)–(v)* all hold, we derive the following two contradictory assertions:

$$\exists d \in \big[\delta(\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})\big] \setminus \{j\} :\ s(\mathcal{V}_{i^*-1},\, d \mid p) < s(\mathcal{V}_{i^*},\, d \mid p), \tag{6.127a}$$

$$\forall d \in \big[\delta(\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})\big] \setminus \{j\} \quad s(\mathcal{V}_{i^*-1},\, d \mid p) > s(\mathcal{V}_{i^*},\, d \mid p). \tag{6.127b}$$

Proof of (6.127a)**.**

Under conditions *(i)–(iii)*, assertion (6.79a) of Proposition 6.3.3 guarantees the existence of $(d,h) \in \mathcal{H}$ satisfying:

$$d \in \delta(\mathcal{H} \setminus \mathcal{U}) \setminus \{j\} \quad \wedge \quad h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) \quad \wedge \quad x_{d,h}^{(i^*-1)} < x_{d,h}^{(i^*)}. \tag{6.128}$$

Let (d,h) be any such pair. To establish (6.127a), we examine three exhaustive cases, each assuming (d,h) lies in one block of the following partition of $\mathcal{H}$:

$$\big\{\mathcal{V}_{i^*-1},\, \mathcal{V}_{i^*} \setminus \mathcal{V}_{i^*-1},\, \mathcal{H} \setminus (\mathcal{V}_{i^*-1} \cup \mathcal{V}_{i^*})\big\}. \tag{6.129}$$

Case 1: $(d,h) \in \mathcal{V}_{i^*-1}$

By Lemma 6.3.7 and (6.128), we have $(d,h) \in \mathcal{V}_{i^*-1} \cap \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H})$. Then,

$$N_{d,h} \overset{(6.4)}{=} x_{d,h}^{(i^*-1)} \overset{(6.128)}{<} x_{d,h}^{(i^*)} \overset{(6.10c)}{\leq} N_{d,h}, \tag{6.130}$$

which leads to a contradiction. Therefore, this case is impossible.

Case 2: $(d,h) \in \mathcal{V}_{i^*} \setminus \mathcal{V}_{i^*-1}$

Recall that $|\mathcal{J}| \overset{(iv)}{\geq} 2$, and consider Proposition 6.3.6 with $\mathcal{A} = \mathcal{J} \setminus \{j\}$. By *(v)*, Assumption 6.3.5 holds for $u = |\mathcal{J}| - 1 = |\mathcal{A}|$; hence, the following implication holds:

$$(d,h) \in \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*})\cap(\mathcal{J}\setminus\{j\})}(\mathcal{H}) \cap \mathcal{V}_{i^*} \quad \Longrightarrow \quad s(\mathcal{V}_{i^*},\, d \mid p) \geq \tfrac{\rho_d}{S_{d,h}}. \tag{6.131}$$

On the other hand, by Proposition 6.3.4 with $\mathcal{A} = \mathcal{J} \setminus \{j\}$, we have

$$(d,h) \in \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*-1})\cap(\mathcal{J}\setminus\{j\})}(\mathcal{H}) \setminus \mathcal{V}_{i^*-1} \quad \Longrightarrow \quad s(\mathcal{V}_{i^*-1},\, d \mid p) < \tfrac{\rho_d}{S_{d,h}}. \tag{6.132}$$

To verify the premises of these implications, observe that:

$$\mathcal{J} \setminus \{j\} \overset{[1]}{\subset} \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}), \tag{6.133a}$$

$$(d,h) \overset{(6.122)}{\in} \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}), \tag{6.133b}$$

$$(d,h) \overset{\text{Case 2}}{\in} \mathcal{V}_{i^*}, \tag{6.133c}$$

$$(d,h) \overset{\text{Case 2}}{\in} \mathcal{H} \setminus \mathcal{V}_{i^*-1}, \tag{6.133d}$$

$$(d,h) \overset{(6.133d)}{\in} \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*-1})}(\mathcal{H}), \tag{6.133e}$$

where [1] follows from (6.62), whose premise is satisfied because $\mathcal{J} \ni j \overset{(ii)}{\in} \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})$. Then, the premise of (6.131) follows from (6.133a)–(6.133c), while the premise of (6.132) follows from (6.133e), (6.133b), and (6.133d) via (2.17e).

Combining the resulting inequalities yields

$$s(\mathcal{V}_{i^*-1},\, d \mid p) < s(\mathcal{V}_{i^*},\, d \mid p). \tag{6.134}$$

Since

$$d \overset{\substack{(6.133b)\\(6.133a)}}{\in} \delta(\mathcal{H}\setminus\mathcal{V}_{i^*}) \quad \text{and} \quad d \overset{(6.133e)}{\in} \delta(\mathcal{H}\setminus\mathcal{V}_{i^*-1}) \quad \text{and} \quad d \overset{(6.128)}{\neq} j, \tag{6.135}$$

this establishes (6.127a).

Case 3: $(d,h) \in \mathcal{H} \setminus (\mathcal{V}_{i^*-1} \cup \mathcal{V}_{i^*})$

By basic set algebra, $\mathcal{H} \setminus (\mathcal{V}_{i^*-1} \cup \mathcal{V}_{i^*}) = (\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap (\mathcal{H} \setminus \mathcal{V}_{i^*})$. Then, applying (6.4) for $i \in \{i^*-1,\, i^*\}$, we obtain:

$$\begin{aligned} x_{d,h}^{(i^*-1)} &= s(\mathcal{V}_{i^*-1},\, d \mid p)\, A_{d,h}(p), \\ x_{d,h}^{(i^*)} &= s(\mathcal{V}_{i^*},\, d \mid p)\, A_{d,h}(p). \end{aligned} \tag{6.136}$$

Substituting these into the inequality in (6.128) yields:

$$s(\mathcal{V}_{i^*-1},\, d \mid p) < s(\mathcal{V}_{i^*},\, d \mid p). \tag{6.137}$$

Since $(d,h) \in (\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap (\mathcal{H} \setminus \mathcal{V}_{i^*})$ and $d \overset{(6.128)}{\neq} j$, this concludes the proof of (6.127a).

In summary, of the three exhaustive cases considered here, Case 1 is impossible, while each of the remaining two yields (6.127a).

Proof of (6.127b)**.**
Let $\mathcal{O} := \big[\delta(\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})\big] \setminus \{j\}$. Under conditions *(i)–(iii)*, assertion (6.79b) of Proposition 6.3.3 ensures that:

$$\forall d \in \mathcal{O} \quad \exists h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) : \; x_{d,h}^{(i^*-1)} > x_{d,h}^{(i^*)}. \tag{6.138}$$

Suppose $\mathcal{O} \neq \emptyset$. Fix an arbitrary $d \in \mathcal{O}$ and let $h \in \eta_d(\mathcal{H} \setminus \mathcal{U})$ be the corresponding element satisfying the inequality in (6.138).

To establish (6.127b), we examine three exhaustive cases, each assuming (d,h) lies in one block of the following partition of $\mathcal{H}$:

$$\big\{\mathcal{V}_{i^*},\, \mathcal{V}_{i^*-1} \setminus \mathcal{V}_{i^*},\, \mathcal{H} \setminus (\mathcal{V}_{i^*-1} \cup \mathcal{V}_{i^*})\big\}. \tag{6.139}$$

Case 1: $(d,h) \in \mathcal{V}_{i^*}$

By Lemma 6.3.7 and (6.138), we have $(d,h) \in \mathcal{V}_{i^*} \cap \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H})$. Then:

$$N_{d,h} \overset{(6.4)}{=} x_{d,h}^{(i^*)} \overset{(6.138)}{<} x_{d,h}^{(i^*-1)} \overset{(6.10c)}{\leq} N_{d,h}, \tag{6.140}$$

which leads to a contradiction. Therefore, this case is impossible.

Case 2: $(d,h) \in \mathcal{V}_{i^*-1} \setminus \mathcal{V}_{i^*}$

Recall that $|\mathcal{J}| \overset{(iv)}{\geq} 2$, and consider Proposition 6.3.6 with $\mathcal{A} = \mathcal{J} \setminus \{j\}$. By *(v)*, Assumption 6.3.5 holds for $u = |\mathcal{J}| - 1 = |\mathcal{A}|$; hence, the following implication holds:

$$(d,h) \in \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*-1})\cap(\mathcal{J}\setminus\{j\})}(\mathcal{H}) \cap \mathcal{V}_{i^*-1} \quad \Longrightarrow \quad s(\mathcal{V}_{i^*-1},\, d \mid p) \geq \tfrac{\rho_d}{S_{d,h}}. \tag{6.141}$$

On the other hand, by Proposition 6.3.4 with $\mathcal{A} = \mathcal{J} \setminus \{j\}$, we have

$$(d,h) \in \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*})\cap(\mathcal{J}\setminus\{j\})}(\mathcal{H}) \setminus \mathcal{V}_{i^*} \quad \Longrightarrow \quad s(\mathcal{V}_{i^*},\, d \mid p) < \tfrac{\rho_d}{S_{d,h}}. \tag{6.142}$$

To verify the premises of these implications, observe that:

$$(d,h) \overset{d\in\mathcal{O}}{\in} \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*-1})}(\mathcal{H}), \tag{6.143a}$$

$$(d,h) \overset{d\in\mathcal{O}}{\in} \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*})}(\mathcal{H}), \tag{6.143b}$$

$$(d,h) \overset{(6.122)}{\in} \gamma_{\mathcal{J}\setminus\{j\}}(\mathcal{H}), \tag{6.143c}$$

$$(d,h) \overset{\text{Case 2}}{\in} \mathcal{V}_{i^*-1}, \tag{6.143d}$$

$$(d,h) \overset{\text{Case 2}}{\in} \mathcal{H} \setminus \mathcal{V}_{i^*}. \tag{6.143e}$$

Then, utilizing (2.17e), the premise of (6.141) follows from (6.143a), (6.143c), and (6.143d); while the premise of (6.142) follows from (6.143b), (6.143c), and (6.143e). Combining the resulting inequalities yields

$$s(\mathcal{V}_{i^*-1},\, d \mid p) > s(\mathcal{V}_{i^*},\, d \mid p). \tag{6.144}$$

<u>Case 3: $(d,h) \in \mathcal{H} \setminus (\mathcal{V}_{i^*-1} \cup \mathcal{V}_{i^*})$</u>

By basic set algebra, $\mathcal{H} \setminus (\mathcal{V}_{i^*-1} \cup \mathcal{V}_{i^*}) = (\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap (\mathcal{H} \setminus \mathcal{V}_{i^*})$. Then, applying (6.4) for $i \in \{i^*-1,\, i^*\}$, we obtain:

$$\begin{aligned} x_{d,h}^{(i^*-1)} &= s(\mathcal{V}_{i^*-1},\, d \mid p)\, A_{d,h}(p), \\ x_{d,h}^{(i^*)} &= s(\mathcal{V}_{i^*},\, d \mid p)\, A_{d,h}(p). \end{aligned} \tag{6.145}$$

Substituting these into the inequality in (6.138) yields:

$$s(\mathcal{V}_{i^*-1},\, d \mid p) > s(\mathcal{V}_{i^*},\, d \mid p). \tag{6.146}$$

In summary, among the three exhaustive cases considered here, Case 1 is impossible, while each of the remaining two yields $s(\mathcal{V}_{i^*-1},\, d \mid p) > s(\mathcal{V}_{i^*},\, d \mid p)$. Since d was chosen arbitrarily from $\mathcal{O}$, this establishes the desired result (6.127b).

To conclude the proof of the proposition, observe that assertions (6.127a) and (6.127b) are mutually contradictory. Since both follow from conditions *(i)–(v)*, these conditions cannot all hold simultaneously, completing the proof by contradiction. □

Theorem 6.3.9 (Monotonicity of s for $i = i^*$)**.** *Assumption 6.3.5 holds for all $u \in \mathbb{N}$.*

Proof. Define the following predicate:

$B(k)$: For the program $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ with any valid input parameter satisfying $|\mathcal{J}| = k$, the following implication holds:

$$\Big(i^* \geq 2 \quad \wedge \quad j \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})\Big) \quad \Longrightarrow \quad s(\mathcal{V}_{i^*-1},\, j \mid p) \leq s(\mathcal{V}_{i^*},\, j \mid p). \tag{6.147}$$

To prove the theorem, it suffices to show that $B(k)$ holds for all $k \in \mathbb{N}$. We establish this using induction, specifically, complete induction. In what follows, when we refer to the context of $B(k)$ for some $k \in \mathbb{N}$, we mean the setting described in the definition of $B(k)$.

Base Case. We aim to prove that $B(1)$ is true. The proof proceeds by contradiction. Suppose, for the sake of contradiction, that $B(1)$ is false. That is, in the context of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program with $|\mathcal{J}| = 1$, we assume:

$$i^* \geq 2 \quad \wedge \quad j \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}), \tag{6.148a}$$

$$s(\mathcal{V}_{i^*-1},\, j \mid p) > s(\mathcal{V}_{i^*},\, j \mid p). \tag{6.148b}$$

Note that by Corollary 6.3.1 and (6.148a), the function values in (6.148b) are well-defined.

Under (6.148) and $\mathcal{J} = \{j\}$, assertion (6.79a) of Proposition 6.3.3 guarantees the existence of $(d, h) \in \mathcal{H}$ satisfying:

$$d \in \delta(\mathcal{H} \setminus \mathcal{U}) \setminus \mathcal{J} \quad \wedge \quad h \in \eta_d(\mathcal{H} \setminus \mathcal{U}) \quad \wedge \quad x_{d,h}^{(i^*-1)} < x_{d,h}^{(i^*)}. \tag{6.149}$$

Let (d, h) be any such pair.

In view of (6.149), it follows from (5.63a) that

$$d \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*-1}) \cap \delta(\mathcal{H} \setminus \mathcal{V}_{i^*}), \tag{6.150a}$$

while from (6.25a) one has

$$\forall w \in \eta_d(\mathcal{H} \setminus \mathcal{U}) \quad x_{d,w}^{(i^*-1)} < x_{d,w}^{(i^*)}. \tag{6.150b}$$

On the other hand, given (6.150a), assertion (6.79b) of Proposition 6.3.3 ensures that

$$\exists z \in \eta_d(\mathcal{H} \setminus \mathcal{U}) : \; x_{d,z}^{(i^*-1)} > x_{d,z}^{(i^*)}. \tag{6.151}$$

Statements (6.150b) and (6.151) are contradictory. Hence the assumption in (6.148) is false, and $B(1)$ holds.

Inductive Step. We aim to prove that for all $k \in \mathbb{N} \setminus \{1\}$,

$$\big(\forall t \in \{1, \ldots, k-1\} \quad B(t) \text{ is true}\big) \quad \Longrightarrow \quad B(k) \text{ is true.} \tag{6.152}$$

The proof proceeds by contradiction. Suppose there exists some $k \in \mathbb{N} \setminus \{1\}$ for which implication (6.152) fails, meaning:

(i) $B(t)$ holds for all $t \in \{1, \ldots, k-1\}$;

and, in the context of the $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ program with $|\mathcal{J}| = k$:

(ii) $i^* \geq 2$,

(iii) $j \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})$,

(iv) $s(\mathcal{V}_{i^*-1}, j \mid p) > s(\mathcal{V}_{i^*}, j \mid p)$.

By the same argument as in the base case, the function values in (iv) are well-defined.

Observe that assumption (i) is precisely Assumption 6.3.5 with $u = k - 1 = |\mathcal{J}| - 1$. Under this observation, and recalling that $|\mathcal{J}| = k \geq 2$, the set of assumptions (i)–(iv) corresponds exactly to the five conditions of Proposition 6.3.8. Since that proposition asserts that these conditions cannot hold simultaneously, we obtain a contradiction. Therefore, the implication (6.152) holds for all $k \in \mathbb{N} \setminus \{1\}$.

By complete induction, $B(k)$ holds for all $k \in \mathbb{N}$, which completes the proof. □

Theorem 6.3.10. *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. If* $\mathcal{J} = \delta(\mathcal{H})$ *and* $\mathcal{V}_{i^*} \subsetneq \mathcal{H}$*, then, for all* $(d, h) \in \mathcal{H}$*,*

$$(d, h) \in \mathcal{V}_{i^*} \quad \Longleftrightarrow \quad s(\mathcal{V}_{i^*}, \boldsymbol{v}, d \mid p) \geq \tfrac{\rho_d}{S_{d,h}}, \tag{6.153}$$

where $(\lambda, \boldsymbol{v}) := \mathrm{Eigen}(\mathcal{V}_{i^*} \mid p)$.

Proof. First, observe that

$$\delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \overset{\text{Th. 6.2.5}}{=} \delta(\mathcal{H}) = \mathcal{J}, \tag{6.154}$$

which implies

$$\gamma_{\delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \cap \mathcal{J}}(\mathcal{H}) = \mathcal{H}. \tag{6.155}$$

(⇒) By Theorem 6.3.9, Assumption 6.3.5 holds for all $u \in \mathbb{N}$. Hence, by Proposition 6.3.6 for the set $\mathcal{A} = \mathcal{J}$, the following implication is valid for all $(d, h) \in \gamma_{\delta(\mathcal{H} \setminus \mathcal{V}_{i^*})}(\mathcal{H})$:

$$(d, h) \in \gamma_{\delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) \cap \mathcal{J}}(\mathcal{H}) \cap \mathcal{V}_{i^*} \quad \Longrightarrow \quad s(\mathcal{V}_{i^*}, \boldsymbol{v}, d \mid p) \geq \tfrac{\rho_d}{S_{d,h}}. \tag{6.156}$$

This completes the necessity part, after referring to (6.155), and noting that $\gamma_{\delta(\mathcal{H} \setminus \mathcal{V}_{i^*})}(\mathcal{H}) \overset{(6.154)}{=} \mathcal{H}$.

($\Leftarrow$) By Proposition 6.3.4 with $\mathcal{A} = \mathcal{J}$, and using the law of contraposition, the following implication holds for all $(d, h) \in \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*})}(\mathcal{H})$:

$$s(\mathcal{V}_{i^*},\, \boldsymbol{v},\, d \mid p) \geq \tfrac{\rho_d}{S_{d,h}} \quad \Longrightarrow \quad (d,h) \notin \gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*})\cap\mathcal{J}}(\mathcal{H}) \setminus \mathcal{V}_{i^*} \overset{(6.155)}{=} \mathcal{H} \setminus \mathcal{V}_{i^*}. \tag{6.157}$$

Note that by Corollary 6.3.1, $s(\mathcal{V}_{i^*},\, \boldsymbol{v},\, d \mid p)$ is well-defined for all $d \in \delta(\mathcal{H} \setminus \mathcal{V}_{i^*})$. This completes the sufficiency part, after noting that $\gamma_{\delta(\mathcal{H}\setminus\mathcal{V}_{i^*})}(\mathcal{H}) \overset{(6.154)}{=} \mathcal{H}$.

□

6.4. Partial and Total Correctness

Theorem 6.4.1 (Admissibility of $\mathcal{V}_i$)**.** *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. For any* $i \in \{1, \ldots, i^*\}$*, the following implications hold:*

$$\mathcal{V}_i \subsetneq \mathcal{H} \quad \Longleftrightarrow \quad n > \sum_{(d,h)\in\mathcal{V}_i} N_{d,h}, \tag{6.158a}$$

$$\mathcal{V}_i = \mathcal{H} \quad \Longrightarrow \quad n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}. \tag{6.158b}$$

Proof. Let $i \in \{1, \ldots, i^*\}$.

The backward implication ($\Leftarrow$) in (6.158a) follows immediately from Remark 2.2.1 after referring to Definition 2.2.3.

By Lemma 6.1.1, the parameters p and $\mathcal{V}_i$ satisfy the input requirements of `DCA`; hence, the following assertion must hold:

$$\Big(n > \sum_{(d,h)\in\mathcal{V}_i} N_{d,h}\Big) \;\vee\; \Big(\mathcal{V}_i = \mathcal{H} \;\wedge\; n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big). \tag{6.159}$$

Given that $n > \sum_{(d,h)\in\mathcal{H}} N_{d,h}$ is impossible by (2.4e), the forward implications in (6.158) follow directly from (6.159). □

Theorem 6.4.2 (Partial correctness)**.** *Consider the* $\texttt{RDCA}(p, \mathcal{U}, \mathcal{J})$ *program. If* $\mathcal{J} = \delta(\mathcal{H})$ *then the point* $(T, \boldsymbol{x})$ *returned by this program is of the form:*

$$T = \begin{cases} 0, & \text{if } \mathcal{V}_{i^*} = \mathcal{H} \\ \lambda, & \text{if } \mathcal{V}_{i^*} \subsetneq \mathcal{H}, \end{cases} \qquad x_{d,h} = \begin{cases} N_{d,h}, & (d,h) \in \mathcal{V}_{i^*} \\ s(\mathcal{V}_{i^*},\, \boldsymbol{v},\, d \mid p)\, A_{d,h}(p), & (d,h) \in \mathcal{H} \setminus \mathcal{V}_{i^*}, \end{cases} \tag{6.160}$$

where $\boldsymbol{x} = \big(x_{d,h},\, (d,h) \in \mathcal{H}\big)$*, and the following conditions hold:*

If $\mathcal{V}_{i^*} \subsetneq \mathcal{H}$ then:

$$n > \sum_{(d,h)\in\mathcal{V}_{i^*}} N_{d,h}, \tag{6.161a}$$

$$(\lambda,\, \boldsymbol{v}) := \mathrm{Eigen}(\mathcal{V}_{i^*} \mid p), \tag{6.161b}$$

$$\delta(\mathcal{H} \setminus \mathcal{V}_{i^*}) = \delta(\mathcal{H}), \tag{6.161c}$$

$$\big(\forall (d,h) \in \mathcal{H}\big) \Big((d,h) \in \mathcal{V}_{i^*} \quad \Leftrightarrow \quad s(\mathcal{V}_{i^*},\, \boldsymbol{v},\, d \mid p) \geq \tfrac{\rho_d}{S_{d,h}} \Big). \tag{6.161d}$$

If $\mathcal{V}_{i^*} = \mathcal{H}$ then:

$$n = \sum_{(d,h)\in\mathcal{H}} N_{d,h}. \tag{6.162}$$

Proof. According to Definition 5.5.1, $\mathcal{V}_{i^*} \subseteq \mathcal{H}$. The required properties are ensured as follows:

- Conditions (6.160) and (6.161b) follow from Theorem 6.1.2 for $i = i^*$.
- Conditions (6.161a) and (6.162) follow from Theorem 6.4.1 for $i = i^*$.
- Condition (6.161c) is ensured by Theorem 6.2.5.
- Condition (6.161d) is ensured by Theorem 6.3.10.

□

Theorem 6.4.3 (Total correctness)**.** *Let $p \in \mathcal{P}$. The* `RDCA`$(p,\, \mathcal{U},\, \mathcal{J})$ *program with $\mathcal{U} = \emptyset$ and $\mathcal{J} = \delta(\mathcal{H})$ terminates and returns a pair $(T,\, \boldsymbol{x})$ that solves the* CPDA*(p) problem.*

Proof. Algorithm termination is guaranteed by Proposition 5.2.2, which establishes that $i^* < \infty$. Upon termination, Theorem 6.4.2 ensures that the returned point $(T,\, \boldsymbol{x})$ satisfies the optimality conditions (3.61)–(3.63) of Theorem 3.2.8, with the assignments $(T^*,\, \boldsymbol{x}^*) = (T,\, \boldsymbol{x})$ and $\mathcal{U}^* = \mathcal{V}_{i^*}$.

According to Theorem 3.2.8, any point of this form is a solution to the CPDA(p) problem. This completes the proof of Theorem 6.4.3.

We further note that in Theorem 3.2.8, the distinction between *Case 1* and *Case 2* is defined by the conditions (3.62)–(3.63). Remark 3.2.5 ensures that the requirements of *Case 1* (specifically, (3.62a)) imply $\mathcal{U}^* \subsetneq \mathcal{H}$, whereas *Case 2* corresponds to the situation where $\mathcal{U}^* = \mathcal{H}$. Thus, while Theorem 3.2.8 does not explicitly partition the conditions for $\mathcal{U}^*$ by the equality $\mathcal{U}^* = \mathcal{H}$, this dichotomy is implicitly enforced.

Consequently, the case distinction in Theorem 6.4.2 – based on whether $\mathcal{V}_{i^*} \subsetneq \mathcal{H}$ or $\mathcal{V}_{i^*} = \mathcal{H}$ – is fully consistent with the two optimality cases defined in Theorem 3.2.8. □

Chapter 7

Discussion of the Results and Future Work

7.1. Summary of Results

In this thesis, we investigated the optimum sample allocation problem, formulated as CPDA(p) for a given population-total sample size model $p \in \mathcal{P}$ (Def. 2.2.2). The problem was posed as a constrained optimization problem, with explicitly defined variables, feasibility constraints, and structural assumptions. We established the existence of an optimal solution and derived sufficient optimality conditions (Th. 3.2.8) using the Karush-Kuhn-Tucker Theorem F.2.1.

The primary contribution of this work is the `RDCA` algorithm, designed to solve the CPDA(p) problem. This algorithm extends the existing `DCA`(p, $\emptyset$) method, which addresses Problem G.2 – a relaxed version of CPDA(p) omitting the inequality constraints (3.2c). By leveraging the optimality conditions stated in Theorem 3.2.8, `RDCA` identifies the set $\mathcal{U}^* \subseteq \mathcal{H}$ of *take-max* strata corresponding to the optimal solution. Once this set is determined, the analytical formula for the solution follows directly from the optimality conditions.

Because the `RDCA` algorithm is specifically tailored to the structural properties of the optimal solution, it avoids the computational drawbacks common to general-purpose nonlinear optimization solvers, such as numerical instability, sensitivity to initial values, or failure to converge to a global optimum (see Sec. 1.3.2).

Furthermore, as an integral part of this research, we provided a robust implementation of the `RDCA` algorithm in the R programming language, as detailed in Section 4.3.

7.2. Directions for Future Research

There are two main directions for future research. The first concerns the development of new algorithms for solving the CPDA problem, while the second focuses on extensions and generalizations of the CPDA. We briefly discuss both directions below.

7.2.1. New algorithms

The `RDCA` algorithm for the CPDA problem is based on a recursive structure in which a loop invokes the algorithm itself at each iteration. This pattern can lead to rapid growth in computational complexity, with the total number of recursive calls increasing roughly exponentially. In practical applications where the number of domains and strata is relatively small (e.g., fewer than 20 domains and 50 strata per domain) and the total sample size is moderate, this remains manageable with modern computing resources. However, as the number of domains increases, the algorithm remains efficient only when the total sample size is relatively small. The computation time grows with the number of domains, the number of strata within each domain, and the total sample size. Therefore, it would be desirable to develop a more efficient (i.e., sub-exponential time) method for identifying the optimal set of *take-max* strata.

A possible – though modest – direction for improving efficiency is to introduce a preprocessing step in which, for each domain $d \in \delta(\mathcal{H})$, the strata are sorted in decreasing order of the ratios $\left(\frac{\rho_d}{S_{d,h}}\right)_{h \in \eta_d(\mathcal{H})}$. This allows the algorithm, when constructing the set $\mathcal{Y}$ on line 10, to examine only an initial segment of the sorted strata for each domain, rather than checking all strata in arbitrary order.

Such an approach is analogous to the method proposed by Stenger and Gabler [41] for solving the classical optimum allocation problem, i.e., the CPDA(p) problem with a single domain ($|\delta(\mathcal{H})| = 1$). Although this modification does not change the algorithm's overall computational complexity, it could offer modest practical improvements in efficiency.

7.2.2. Generalizations of the CPDA problem

Several possible extensions of the CPDA problem are of interest from an applied perspective.

One such extension is to include lower bounds on sample sizes in strata, i.e., constraints of the form $x_{d,h} \geq m_{d,h}$ for all $(d, h) \in \mathcal{H}$, where $m_{d,h} \in (0, N_{d,h}]$. In this context, it may

be useful to refer to Wesołowski et al. [47] and Münnich, Sachs and Wagner [31], who provide algorithms for solving such an extended CPDA(p) problem in the case of a single domain ($|\delta(\mathcal{H})| = 1$).

Another natural extension is to impose integrality constraints on the allocation variables, i.e., $x_{d,h} \in \mathbb{N}$ for all $(d, h) \in \mathcal{H}$, resulting in an integer optimum allocation problem. Wright [49, 50] discuss such an extended CPDA problem for the single-domain case in the presence of box constraints on the sample sizes within strata.

It would also be interesting to investigate multivariate extensions of the CPDA problem, since in many applications one seeks an allocation that is jointly optimal with respect to several study variables. The multivariate version of the CPDA problem for the single-domain case is considered in de Moura Brito, do Nascimento Silva, Silva Semaan and Maculan [12], where the authors also include constant lower-bound constraints on the stratum sample sizes. The proposed solution procedure employs a binary integer programming approach; see also Brito, Silva and Veiga [6] for an R implementation of this method. A general multiobjective perspective on optimum allocation in survey sampling is provided by Willems [48], who formulates the problem as a vector-valued optimization task accounting for potentially conflicting study variables and practical constraints. The framework is sufficiently general to encompass allocation problems closely related to the CPDA formulation, which can be viewed as a particular instance within this broader class.

Finally, it would be worthwhile to explore multi-stage sampling extensions of the CPDA problem. For the relaxed version of CPDA without inequality constraints (3.2c), two-stage sampling schemes have been examined by Wesołowski and Wieczorkowski [45] and earlier by Kozak, Zieliński and Singh [24], including applications to agricultural surveys.

In addition to these extensions motivated by practical considerations, further theoretical work could investigate whether the sufficient optimality conditions established in Theorem 3.2.8 of Chapter 3 are also *necessary*. Establishing necessary and sufficient conditions would provide a complete optimality characterization of the CPDA problem and deepen the understanding of its mathematical structure.

Appendix A

List of Symbols

$\mathbb{N}_0$	set of natural numbers including zero, i.e., $\{0, 1, 2, \dots\}$
$\mathbb{N}$	set of natural numbers, i.e., $\{1, 2, \dots\}$
$\mathbb{N}^n$	set of n-tuples $x = (x_1, \dots, x_n)$ with $x_i \in \mathbb{N}$, i.e., $\bigtimes_n \mathbb{N}$
$\mathbb{Z}$	set of integers, i.e., $\{\dots, -2, -1, 0, 1, 2, \dots\}$
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$	set of positive real numbers, i.e., $(0, +\infty)$
$\mathbb{R}^n$	set of n-tuples $x = (x_1, \dots, x_n)$ with $x_i \in \mathbb{R}$, i.e., $\bigtimes_n \mathbb{R}$
$\mathbb{R}^n_+$	set of n-tuples $x = (x_1, \dots, x_n)$ with $x_i \in \mathbb{R}_+$, i.e., $\bigtimes_n \mathbb{R}_+$
$\mathbb{R}^{m\times m}$	set of matrices $A = (a_{ij})$ with $a_{ij} \in \mathbb{R}$ for $1 \leq i, j \leq m$
$\mathbb{C}$	set of complex numbers
$\mathbf{0}$	zero vector, i.e., $(0, \dots, 0)$ or $[0, \dots, 0]^\top$ depending on context
$A \times B$	cartesian product of sets A and B, i.e., $\{(a, b)\colon\ a \in A,\, b \in B\}$
$\bigtimes_{i\in I} A_i$	cartesian product of a family of sets $\{A_i\}_{i\in I}$ for indexing set $I \neq \emptyset$, that is, $\bigtimes_{i\in I} A_i := \big\{(a_i,\, i \in I)\colon\ a_i \in A_i \text{ for all } i \in I\big\}$
$\bigtimes_n A$	shorthand for $\bigtimes_{i\in\{1,\dots,n\}} A$
$A^{\complement}$	complement of set A with respect to the implicit universe under consideration
$\lvert A\rvert$	cardinality of set A
$\lVert x\rVert$	Euclidean norm of vector $x \in \mathbb{R}^n$, i.e., $\sqrt{\sum_{i=1}^n x_i^2}$

$x \circ y$	component-wise (Hadamard) multiplication of vectors x and y, i.e., $(x \circ y)_i = x_i y_i$
$\mathbf{A}^\top$	transpose of matrix $\mathbf{A}$
$\mathrm{diag}(\boldsymbol{b})$	diagonal matrix with vector $\boldsymbol{b}$ on the diagonal
$\vee$, $\wedge$	logical OR (disjunction), logical AND (conjunction)
$\implies$, $\iff$	logical implication, logical equivalence
$\mathbb{E}[X]$	expected value of random variable X
$(\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n)$	population-total sample size model (see Def. 2.2.1)
$\mathcal{P}$	family of population-total sample size models (see Def. 2.2.2)
δ, $\eta_\cdot$, $\gamma_\cdot$	index-extractor auxiliary functions (see Def. 2.3.1)
$A_{d,h}$, c_d	model-parameter auxiliary functions (see Def. 2.3.2)
$\mathcal{F}_p$	family of admissible sets for model $p \in \mathcal{P}$ (see Def. 2.2.3)
$\mathbf{D}_{\mathcal{A}\|p}$	population-total sample size matrix for $\mathcal{A} \in \mathcal{F}_p$, $p \in \mathcal{P}$ (see Def. 3.2.1)
Eigen	eigenpair operator (see Def. 3.2.2)
s	function (see Def. 3.2.3)
$\mathcal{V}_i$	take-max strata sets $\mathcal{V}_i$ (see Def. 5.5.1 or Rem. 5.5.1)

Appendix B

List of Abbreviations

Abbreviation	Reference	Description
CRAN	R Core Team [35]	Comprehensive R Archive Network
CPDA	Problem 3.1.1	Controlled-Precision Domain Allocation problem
DCA	Algorithm 2	Domain-Controlled Allocation algorithm
KKT	Theorem F.2.1	Karush-Kuhn-Tucker conditions
LBRP	Definition 5.4.1	Last-Branch Recursion Path
NLP		Nonlinear programming
RDCA	Algorithm 1	Recursive Domain-Controlled Allocation algorithm
REL-CPDA	Problem 3.2.1	Relaxed Controlled-Precision Domain Allocation problem
RNA	Wesołowski et al. [46]	Recursive Neyman Algorithm
SI	Särndal et al. [42, Sec. 3.3, p. 66]	Simple random sampling without replacement
ST	Särndal et al. [42, Sec. 3.7, p. 100]	Stratified sampling
STSI	same as above	Stratified simple random sampling without replacement

Appendix C

Perron-Frobenius Theory

The following properties are classical statements from the **Perron-Frobenius theorem**.

Theorem C.1 (Perron-Frobenius)**.** *Let* $\mathbf{A}$ *be a real square matrix with all entries strictly positive. Then the following statements hold:*

1. *There exists a real, strictly positive and simple eigenvalue of* $\mathbf{A}$*, denoted by* r *and known as the* ***Perron root****, such that* $r > |\lambda|$ *for any other eigenvalue* λ *of* $\mathbf{A}$*.*
2. *There exists a real eigenvector* $\boldsymbol{v}$ *associated with eigenvalue* r *such that* $\boldsymbol{v} > \mathbf{0}$*. This vector, called the* ***Perron vector****, is unique up to positive scaling.*
3. *Any eigenvector of* $\mathbf{A}$ *that is not a positive scalar multiple of* $\boldsymbol{v}$ *must have at least one negative component.*

For a proof of these statements, see, e.g., Meyer [28, Sec. 8.2, p. 663], Gentle [15, Sec. 8.7.2, p. 373], or Kato [20, Chapter I, Par. 7, Sec. 2, p. 66].

Appendix D

Convex Sets and Convex Functions

Although the general definition of a convex function (together with the corresponding theorems) applies to functions $f\colon \mathbb{R}^n \to [-\infty, +\infty]$, in the following definitions we restrict our attention to finite-valued functions $f\colon \mathbb{R}^n \to \mathbb{R}$.

Definition D.1. Let $X \subset \mathbb{R}^n$ and let $f\colon X \to \mathbb{R}$. The **epigraph** of f is the subset of $\mathbb{R}^{n+1}$ defined by

$$\operatorname{epi} f := \{(x, \mu) \in X \times \mathbb{R}\colon\ f(x) \leq \mu\}. \tag{D.1}$$

Definition D.2. A subset $X \subset \mathbb{R}^n$ is called **convex** if for all $x_1,\, x_2 \in X$,

$$(1-\lambda)\, x_1 + \lambda\, x_2 \in X, \qquad \forall \lambda \in [0,\, 1]. \tag{D.2}$$

Definition D.3. Let $X \subset \mathbb{R}^n$ and let $f\colon X \to \mathbb{R}$. The function f is called **convex** if its epigraph is a convex subset of $\mathbb{R}^{n+1}$.

Proposition D.1. *Let $f_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in \{1, \ldots, m\}$, be convex functions. Then the following functions are convex:*

- *nonnegative weighted sums $\sum_{i=1}^m w_i f_i + b$, where $w_i \geq 0$ for all i and $b \in \mathbb{R}$;*
- *the pointwise maximum function $g\colon \mathbb{R}^n \to \mathbb{R}$, defined by*

$$g(x) := \max\bigl(f_1(x), \ldots, f_m(x)\bigr), \qquad x \in \mathbb{R}^n. \tag{D.3}$$

Proof. See, e.g., Bertsekas [4, Prop. B.2, p. 785] or Mordukhovich and Nam [29, Th. 1.38, p. 14]. □

Appendix E

Selected Results in Topology

This appendix provides a brief review of the topological definitions and results required for the analysis in Chapter 3. For our purposes, it is sufficient to consider these concepts within the framework of metric spaces.

When referring to a metric space $(M,\, d)$, we generally denote it simply by M whenever the metric is clear from the context or immaterial to the results. Unless otherwise stated, $\mathbb{R}^n$ is equipped with the standard Euclidean metric $d(x,\, y) := \|x - y\|$ for $x,\, y \in \mathbb{R}^n$.

E.1. Preliminaries on Metric Topology

Definition E.1.1. Let $(M,\, d)$ be a metric space, and let $X \subset M$. Define the metric d_X on X by

$$d_X(x_1,\, x_2) := d(x_1,\, x_2), \qquad x_1,\, x_2 \in X. \tag{E.1}$$

Then $(X,\, d_X)$ is a metric space, called a **subspace** of $(M,\, d)$.

Definition E.1.2. Let $(M,\, d)$ be a metric space. The **open ball** centered at $x \in M$ with radius $r > 0$ is the set

$$B_M(x,\, r) := \{y \in M :\ d(x,\, y) < r\}. \tag{E.2}$$

Definition E.1.3. Let $(M,\, d)$ be a metric space. A subset $X \subset M$ is called **open** in $(M,\, d)$ if

$$\forall x \in X \quad \exists r > 0 :\ B_M(x,\, r) \subset X, \tag{E.3}$$

where B_M denotes the open ball in metric space $(M,\, d)$ (see Def. E.1.2).

Definition E.1.4. Let $(M,\, d)$ be a metric space. A subset $X \subset M$ is called **closed** in $(M,\, d)$ if its complement $M \setminus X$ is open in $(M,\, d)$.

Theorem E.1.1. *Let M be a metric space. Then:*

1. *The union of any family of open sets in M is open in M.*
2. *The intersection of any finite family of open sets in M is open in M.*

Proof. See Singh [39, Th. 1.1.5, p. 4]. □

Theorem E.1.2. *Let M be a metric space. Then:*

1. *The intersection of any family of closed sets in M is closed in M.*
2. *The union of any finite family of closed sets in M is closed in M.*

Proof. The results for closed sets follow immediately from Theorem E.1.1 by applying De Morgan's laws to the respective complements. □

Theorem E.1.3. *Let M_1 be a subspace of a metric space M. Then, a subset X of M_1 is:*

1. *Open in M_1 if and only if $X = M_1 \cap U$ for some set U open in M.*
2. *Closed in M_1 if and only if $X = M_1 \cap V$ for some set V closed in M.*

Proof. See Sohrab [40, Th. 5.2.5, p. 187]. □

Theorem E.1.4. *Let M_1 be a subspace of a metric space M. If a subset $X \subset M_1$ is closed in M_1 and M_1 is closed in M, then X is closed in M.*

Proof. Since X is closed in M_1, by Theorem E.1.3, there exists a set V closed in M such that $X = M_1 \cap V$. As both M_1 and V are closed in M, their intersection X is also closed in M (see Th. E.1.2).

See also Munkres [30, Lem. 16.2, p. 89] for the corresponding result for open sets in topological spaces, or Rudin [37, Th. 2.30, p. 36]. □

Definition E.1.5. Let (M, d) be a metric space. A subset $X \subset M$ is **bounded** in (M, d) if there exists $r \geq 0$ such that $d(x_1, x_2) \leq r$ for all $x_1, x_2 \in X$.

Theorem E.1.5. *Let $(\mathbb{R}^n, d)$ be a Euclidean metric space. A subset $X \subset \mathbb{R}^n$ is bounded in $(\mathbb{R}^n, d)$ if there exists $r \geq 0$ such that*

$$\|x\| \leq r, \qquad \forall x \in X. \tag{E.4}$$

Proof. Suppose such r exists. Then, for any $x_1, x_2 \in X$, the triangle inequality implies

$$d(x_1, x_2) = \|x_1 - x_2\| \leq \|x_1\| + \|x_2\| \leq r + r = 2\,r \geq 0. \tag{E.5}$$

□

E.2. Compact Sets

Definition E.2.1. Let M be a metric space. An **open cover** of a set $X \subset M$ is a collection $\{G_\lambda\colon\ \lambda \in \Lambda\}$ of open subsets of M such that

$$X \subset \bigcup_{\lambda \in \Lambda} G_\lambda. \tag{E.6}$$

Definition E.2.2. Let M be a metric space. A subset $X \subset M$ is **compact** if for every open cover $\{G_\lambda\colon\ \lambda \in \Lambda\}$ of X, there exists a finite subcollection $\{G_{\lambda_1}, \ldots, G_{\lambda_n}\}$ such that

$$X \subset \bigcup_{i=1}^{n} G_{\lambda_i}. \tag{E.7}$$

Theorem E.2.1 (Heine-Borel Theorem)**.** *In the Euclidean metric space* $(\mathbb{R}^n, d)$*, a subset* $X \subset \mathbb{R}^n$ *is compact if and only if it is closed and bounded in* $(\mathbb{R}^n, d)$*.*

Proof. See Sohrab [40, Cor. 5.6.38, p. 225]. □

E.3. Connected Sets

Definition E.3.1. A metric space (M, d) is **connected** if there do not exist two disjoint, nonempty sets $U, V \subset M$ that are open in (M, d) such that $M = U \cup V$.

Definition E.3.2. Let X be a subset of a metric space (M, d). We say that X is **connected** if the metric subspace (X, d_X) is connected, where d_X is the induced metric on X.

Theorem E.3.1. *In the Euclidean metric space* $(\mathbb{R}, d)$*, a subset* $X \subset \mathbb{R}$ *is connected if and only if it is an interval.*

Proof. See, e.g., Pons [34, Th. 4.3.7, p. 162] or Rudin [37, Th. 2.47, p. 42]. □

Theorem E.3.2. *Let* $X \subset \mathbb{R}^n$ *be a convex set. Then the metric subspace* (X, d_X) *is connected, where* d_X *is the induced metric on* X*.*

Proof. See Malla and Bajracharya [27, Th. 1, p. 26] and Rudin [37, Ex. 21.(c), p. 45]. □

For more details on the theory of connected sets, see Rudin [37, Ch. 2, p. 42; Ch. 4, p. 93], Sohrab [40, Sec. 5.7, p. 226], or Munkres [30, Ch. 3, p. 145].

E.4. Continuity

Theorem E.4.1. *Let M_1 and M_2 be metric spaces. A function $f\colon M_1 \to M_2$ is continuous if and only if, for every set $Y \subset M_2$ that is open in M_2, the preimage $f^{-1}(Y)$ is open in M_1.*

Proof. See Singh [39, Th. 1.1.6, p. 4]. □

Theorem E.4.2. *Let M_1 and M_2 be metric spaces. A function $f\colon M_1 \to M_2$ is continuous if and only if, for every set $Y \subset M_2$ that is closed in M_2, the preimage $f^{-1}(Y)$ is closed in M_1.*

Proof.

(⇒) Suppose f is continuous and let $Y \subset M_2$ be closed in M_2. By the properties of preimages, we have

$$M_1 \setminus f^{-1}(Y) = f^{-1}(M_2 \setminus Y). \tag{E.8}$$

Since Y is closed, $M_2 \setminus Y$ is open in M_2. By Theorem E.4.1, $f^{-1}(M_2 \setminus Y)$ is open in M_1. Thus, its complement $f^{-1}(Y)$ is closed in M_1.

(⇐) Suppose the preimage of every closed set in M_2 is closed in M_1. For any open set $U \subseteq M_2$, its complement $M_2 \setminus U$ is closed. By our assumption, the set

$$f^{-1}(M_2 \setminus U) = M_1 \setminus f^{-1}(U) \tag{E.9}$$

is closed in M_1. Consequently, $f^{-1}(U)$ is open in M_1, and thus, f is continuous (Th. E.4.1).

□

Theorem E.4.3. *Let (M_1, d_1) and (M_2, d_2) be metric spaces, and let $f\colon M_1 \to M_2$ be a continuous map. If $X \subset M_1$ is connected in (M_1, d_1), then $f(X)$ is connected in (M_2, d_2).*

Proof. See Sohrab [40, Th. 5.7.22, p. 230], Munkres [30, Th. 23.5, p. 150], or Rudin [37, Th. 4.22, p. 93].

□

Definition E.4.1. Let $X \subset \mathbb{R}^n$. A function $f\colon X \to \mathbb{R}$ is **lower semicontinuous** at $x \in X$ if, for every sequence $\{x_k\} \subset X$ such that $x_k \to x$, we have

$$f(x) \leq \liminf_{k\to\infty} f(x_k). \tag{E.10}$$

Corollary E.4.4. *If a function $f\colon X \to \mathbb{R}$ is continuous at $x \in X$, then it is lower semicontinuous at x.*

Proof. If f is continuous at $x \in X$, then for any sequence $\{x_k\} \subset X$ such that $x_k \to x$,

$$\liminf_{k\to\infty} f(x_k) = \lim_{k\to\infty} f(x_k) = f(x). \tag{E.11}$$

□

Definition E.4.2. Let $X \subset \mathbb{R}^n$. A function $f\colon X \to \mathbb{R}$ is **coercive** if, for every sequence $\{x_k\} \subset X$ such that $\|x_k\| \to \infty$, it follows that

$$\lim_{k\to\infty} f(x_k) = \infty. \tag{E.12}$$

Theorem E.4.5 (Weierstrass Theorem)**.** *Let $\mathbb{R}^n$ be a metric space, let $X \subset \mathbb{R}^n$ be nonempty, and let $f\colon X \to \mathbb{R}$ be lower semicontinuous at all points of X. Assume that at least one of the following conditions holds:*

1. *X is compact in $\mathbb{R}^n$.*
2. *X is closed in $\mathbb{R}^n$ and f is coercive.*
3. *There exists $\gamma \in \mathbb{R}$ such that the sublevel set*

$$\{x \in X\colon\ f(x) \le \gamma\} \tag{E.13}$$

is nonempty and compact in $\mathbb{R}^n$.

Then, the set of minimizers of f over X is nonempty and compact in $\mathbb{R}^n$.

Proof. See Bertsekas [4, Proposition A.8 (Weierstrass' Theorem), p. 755] and the accompanying proof. □

E.5. Relative Interior

Definition E.5.1. Let (M, d) be a metric space and let $X \subset M$. The **interior** of X in (M, d) is defined by

$$\operatorname{int}(X) := \{x \in X\colon\ \exists r > 0 \text{ such that } B_M(x, r) \subset X\}, \tag{E.14}$$

where B_M denotes the open ball in (M, d) (see Def. E.1.2).

Definition E.5.2. Let $X \subset \mathbb{R}^n$. The **affine hull** of X, denoted $\operatorname{aff}(X)$, is the set of all affine combinations of points in X, i.e.,

$$\operatorname{aff}(X) := \Big\{ \sum_{i=1}^{k} \theta_i x_i : \quad x_i \in X,\ \theta_i \in \mathbb{R},\ \sum_{i=1}^{k} \theta_i = 1,\ k \in \mathbb{N} \Big\}. \tag{E.15}$$

Definition E.5.3. Let $(\mathbb{R}^n,\, d)$ be a Euclidean metric space. The **relative interior** of a subset $X \subset \mathbb{R}^n$, denoted $\operatorname{relint}(X)$, is the interior of X relative to its affine hull $\operatorname{aff}(X)$, i.e.,

$$\operatorname{relint}(X) := \big\{ x \in X : \ \exists r > 0 \text{ such that } \big(B_{\mathbb{R}^n}(x,\, r) \cap \operatorname{aff}(X) \subset X \big) \big\}, \tag{E.16}$$

where $B_{\mathbb{R}^n}$ denotes the open ball in $(\mathbb{R}^n,\, d)$ (see Def. E.1.2).

Remark E.5.1 (Motivation for relative interior)**.** The concept of relative interior is useful when a set has an empty interior with respect to the ambient space (i.e., the space in which this set is embedded), yet possesses a non-empty interior relative to the lower-dimensional affine subspace in which it lies. Consider the following example:

- Let $X := [0,\, 1] \times \{0\} \subset \mathbb{R}^2$.
- The standard interior of X in $\mathbb{R}^2$ is empty:

$$\operatorname{int}(X) = \emptyset, \tag{E.17}$$

  since every open ball in $\mathbb{R}^2$ around a point of X contains points outside of X.
- The affine hull of X is the x-axis:

$$\operatorname{aff}(X) = \{(x,\, y) \in \mathbb{R}^2 : \ y = 0\}. \tag{E.18}$$

- The relative interior of X in $\operatorname{aff}(X)$ is the open interval along the x-axis:

$$\operatorname{relint}(X) = (0,\, 1) \times \{0\}. \tag{E.19}$$

- Thus, $\operatorname{relint}(X)$ provides a well-defined notion of "interior" for sets whose dimension is lower than that of the ambient space.

For more details, see Boyd and Vandenberghe [5, Sec. 2.1.3 Affine dimension and relative interior, p. 23]

Appendix F

Elements of Mathematical Optimization

This appendix summarizes the foundational definitions and results from mathematical optimization and convex analysis required to establish the optimality conditions for the CPDA problem, as presented in Section 3.2. For a comprehensive treatment of these topics, we refer the reader to standard texts such as Nocedal and Wright [33] and Rockafellar [36].

F.1. The Optimization Problem

This dissertation focuses on **continuous optimization problems** expressed in the standard form as Problem F.1.1.

Problem F.1.1 (Continuous Optimization Problem)**.** The problem is formulated as follows:

$$\underset{x \in D}{\text{minimize}} \quad f(x) \tag{F.1}$$

$$\text{subject to} \quad h_i(x) = 0, \qquad i \in \{1, \ldots, m\}, \tag{F.2a}$$

$$g_j(x) \leq 0, \qquad j \in \{1, \ldots, t\}, \tag{F.2b}$$

where m, $t \in \mathbb{N}_0$, and

$$f \colon D_f \subset \mathbb{R}^n \to \mathbb{R}, \qquad f \in C^1, \tag{F.3}$$

$$h_i \colon D_{h_i} \subset \mathbb{R}^n \to \mathbb{R}, \qquad h_i \in C^1, \quad i \in \{1, \ldots, m\}, \tag{F.4}$$

$$g_j \colon D_{g_j} \subset \mathbb{R}^n \to \mathbb{R}, \qquad g_j \in C^1, \quad j \in \{1, \ldots, t\}, \tag{F.5}$$

$$D_0 := D_f \cap \Big(\bigcap_{i=1}^{m} D_{h_i}\Big) \cap \Big(\bigcap_{j=1}^{t} D_{g_j}\Big), \tag{F.6}$$

$$D \subset D_0. \tag{F.7}$$

The following standard naming conventions apply to Problem F.1.1:

- D is the problem **domain**, i.e., the set of all admissible values of the optimization variable, which may encode additional structural restrictions;
- x is the **optimization variable**;
- f is the **objective function**;
- h_i, $i \in \{1, \ldots, m\}$, and g_j, $j \in \{1, \ldots, t\}$, are the **equality** and **inequality constraint functions**, respectively.

Definition F.1.1. In Problem F.1.1, the set $X \subset D$ defined by

$$X := \{x \in D\colon\ h_i(x) = 0,\ \forall i,\ g_j(x) \leq 0,\ \forall j\} \tag{F.8}$$

is called the **feasible set**.

Definition F.1.2. Problem F.1.1 is said to be **feasible** if its feasible set is nonempty. Otherwise, the problem is called **infeasible**.

Definition F.1.3. Let $X \subseteq D$ be the feasible set of Problem F.1.1. A point $x \in D$ is called **feasible** if $x \in X$. A point $x \in D \setminus X$ is called **infeasible**.

Definition F.1.4. Let x be a feasible point of Problem F.1.1. The **set of active constraints** at x is the set of indices

$$A(x) := \big\{j \in \{1, \ldots, t\}\colon\ g_j(x) = 0\big\}. \tag{F.9}$$

The jth inequality constraint $g_j(x) \leq 0$ is said to be **active** at x if $j \in A(x)$.

Definition F.1.5. Let X be the feasible set of Problem F.1.1. A feasible point $x^* \in X$ is a **global minimizer** (or an **optimal solution**) of Problem F.1.1 if

$$f(x^*) \leq f(x), \qquad \forall x \in X.$$

The point x^* is a **local minimizer** if there exists $\epsilon > 0$ such that

$$f(x^*) \leq f(x), \qquad \forall x \in \{y \in X\colon \|y - x^*\| < \epsilon\}.$$

Definition F.1.6 of a regular point is crucial for Theorem F.2.1, which is one of the main results of this appendix.

Definition F.1.6. Consider Problem F.1.1. A feasible point x is said to be **regular** if the gradients of the equality constraints and the active inequality constraints,

$$\nabla h_1(x),\ \ldots,\ \nabla h_m(x),\ \nabla g_j(x),\ j \in A(x), \tag{F.10}$$

are linearly independent.

F.2. Optimality Conditions and Convex Optimization

Theorem F.2.1 (Karush-Kuhn-Tucker necessary conditions)**.** *Let x^* be a local minimizer of Problem F.1.1 and assume that x^* is regular. Then there exist constants $\lambda_i^* \in \mathbb{R}$, $i \in \{1, \dots, m\}$, and $\mu_j^* \geq 0$, $j \in \{1, \dots, t\}$, called Karush-Kuhn-Tucker multipliers, such that*

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{t} \mu_j^* \nabla g_j(x^*) &= \mathbf{0}, && \textit{(stationarity)} \\ h_i(x^*) &= 0, \quad i \in \{1, \dots, m\}, && \textit{(primal feasibility)} \\ g_j(x^*) &\leq 0, \quad j \in \{1, \dots, t\}, && \textit{(primal feasibility)} \\ \mu_j^* \, g_j(x^*) &= 0, \quad j \in \{1, \dots, t\}. && \textit{(complementary slackness)} \end{aligned} \tag{F.11}$$

Proof. See e.g. Bertsekas [4, Prop. 4.3.1, p. 379] or Avriel [1, Th. 3.4, p. 34]. The latter provides a proof for the Fritz John conditions, which are slightly weaker necessary conditions that do not require the regularity assumption. □

In many applications, the regularity assumption in Theorem F.2.1 is difficult to verify in practice. To address this issue, alternative assumptions known as **constraint qualifications** (CQs) are used to guarantee the existence of KKT multipliers. While numerous CQs exist, we focus on the one most relevant to the optimization problems considered in this work. For a comprehensive graphical illustration of the relationships among various constraint qualifications, see Bertsekas [4, Fig. 4.3.9, p. 407].

Theorem F.2.2 ((Weak) Slater's Constraint Qualification)**.** *Let x^* be a local minimizer of Problem F.1.1 with domain D. Assume that the functions $h_1, \dots, h_m$ are affine and the functions $g_1, \dots, g_t$ are convex. Suppose there exists a point $x \in \operatorname{relint}(D)$ satisfying*

$$\begin{aligned} h_i(x) &= 0, \quad i \in \{1, \dots, m\}, \\ g_j(x) &\leq 0, \quad j \in J, \\ g_j(x) &< 0, \quad j \in \{1, \dots, t\} \setminus J, \end{aligned} \tag{F.12a}$$

where the index set J is defined as

$$J := \big\{ j \in \{1, \dots, t\} \colon \; g_j \textit{ is an affine function} \big\}. \tag{F.12b}$$

Then the KKT conditions (F.11) *hold at x^*. Here,* $\operatorname{relint}(D)$ *denotes the relative interior of D (see Def. E.5.3).*

Proof. See Boyd and Vandenberghe [5, Sec. 5.2.3, p. 226], Hiriart-Urruty and Lemaréchal [18, Th. 2.2.5, p. 310], or Bertsekas [4, Prop. 4.3.9, p. 395] for the proof of (strong) Slater's constraint qualification. □

The conditions in (F.12) are referred to as the **weak (or refined) Slater condition**.

Remark F.2.1. Note that the refined Slater condition (F.12) reduces to standard feasibility when all constraint functions are affine and the problem domain satisfies $D = \text{relint}(D)$.

Below, we present key properties of convex optimization problems, an important subclass of optimization problems. In particular, Theorem F.2.4 shows that, for convex problems satisfying the refined Slater condition, the KKT conditions (F.11) are both necessary and sufficient for optimality.

Definition F.2.1 (Convex Optimization Problem)**.** Problem F.1.1 is called a **convex optimization problem** if the objective function f is convex on D, and the feasible set is convex.

Remark F.2.2. The feasible set in Problem F.1.1 is convex if the functions $h_1, \ldots, h_m$ are affine on D and the functions $g_1, \ldots, g_t$ are convex on D. This follows directly from the fact that the intersection of convex sets is convex.

Theorem F.2.3. *In a convex optimization problem, any local minimizer is also a global minimizer.*

Proof. See, e.g., Boyd and Vandenberghe [5, Sec. 4.2.2, p. 133]. □

Theorem F.2.4 (Karush-Kuhn-Tucker conditions for convex optimization problem)**.** *Consider a convex optimization problem as defined in Definition F.2.1 for which Slater's condition* (F.12) *holds. A point $x^* \in D$ is an optimal solution if and only if there exist multipliers $\lambda_i^* \in \mathbb{R}$, $i \in \{1, \ldots, m\}$, and $\mu_j^* \geq 0$, $j \in \{1, \ldots, t\}$, such that the KKT conditions* (F.11) *are satisfied.*

Proof. Under the refined Slater condition (F.12), any local minimizer x^* of Problem F.1.1 satisfies the KKT conditions (F.11).

Since the problem is convex, any point satisfying the KKT conditions is a global minimizer (see, e.g., Boyd and Vandenberghe [5, Sec. 5.5.3, p. 243]). □

Appendix G

Multi-Domain Optimum Sample Allocation with Controlled-Precision without Upper-Bound Constraints

Wesołowski and Wieczorkowski [45] and Wesołowski [44] studied a relaxed version of the CPDA problem that omits the inequality constraints (3.2c). This relaxed formulation is denoted here as Problem G.1.

Problem G.1. Given $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$,

$$\underset{(T,\boldsymbol{x})\,\in\,\mathbb{R}\times\mathbb{R}_+^{|\mathcal{H}|}}{\text{minimize}} \quad T \tag{G.1}$$

$$\text{subject to} \quad \sum_{(d,h)\in\mathcal{H}} x_{d,h} - n = 0, \tag{G.2a}$$

$$\sum_{h\in\eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) - T = 0, \qquad d \in \delta(\mathcal{H}), \tag{G.2b}$$

$$T > 0, \tag{G.2c}$$

where $(T,\, \boldsymbol{x}) = \big(T,\, (x_{d,h},\ (d,h) \in \mathcal{H})\big)$ is the optimization variable, and the functions δ, η_d, $A_{d,h}$, and c_d are as defined in Definitions 2.3.1 and 2.3.2.

The authors of Wesołowski and Wieczorkowski [45, Th. 2.1, p. 2215] proved the existence and uniqueness of the solution to Problem G.1 provided that $n \in (0,\, n_{max})$, where

$$n_{max} := \sum_{d\in\delta(\mathcal{H})} \frac{\left(\sum_{h\in\eta_d(\mathcal{H})} N_{d,h} S_{d,h}\right)^2}{\sum_{h\in\eta_d(\mathcal{H})} N_{d,h} S_{d,h}^2}. \tag{G.3}$$

They also proposed an algorithm to compute this solution, namely, $\texttt{DCA}(p,\, \emptyset)$. In Chapter 4, we presented a generalized version of `DCA` that serves as the base-case algorithm within the recursive `RDCA` framework. Regarding this implementation, we note the following:

1. The generalized `DCA` accepts any set $\mathcal{U} \subseteq \mathcal{H}$ satisfying the input requirements as its second parameter, allowing it to handle both empty and nonempty sets $\mathcal{U}$ within the recursive structure of `RDCA`.
2. It can be shown that $n_{max} \in \big(0,\, \sum_{(d,h)\in\mathcal{H}} N_{d,h}\big]$.
3. The reason for specifying a weaker requirement for n in `DCA` – specifically $n \in (0,\, \sum N_{d,h}]$ instead of the stricter $n \in (0,\, n_{max})$ – is explained following Remark G.3.

It is of interest to consider how the solution to Problem G.1 changes when the model $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$ is such that $n \in \big[n_{max},\, \sum_{(d,h)\in\mathcal{H}} N_{d,h}\big]$, where n_{max} is defined as in (G.3). In this context, we consider Problem G.2, a relaxed version of Problem G.1 in which constraint (G.2c) is omitted.

Problem G.2. Given $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$,

$$\underset{(T,\,\boldsymbol{x})\,\in\,\mathbb{R}\times\mathbb{R}_+^{|\mathcal{H}|}}{\text{minimize}} \quad T \tag{G.4}$$

$$\text{subject to} \quad \sum_{(d,h)\in\mathcal{H}} x_{d,h} - n = 0, \tag{G.5a}$$

$$\sum_{h\in\eta_d(\mathcal{H})} \frac{[A_{d,h}(p)]^2}{x_{d,h}} - c_d(p) - T = 0, \qquad d \in \delta(\mathcal{H}), \tag{G.5b}$$

where $(T,\, \boldsymbol{x}) = \big(T,\, (x_{d,h},\ (d,h) \in \mathcal{H})\big)$ is the optimization variable, and the functions δ, η_d, $A_{d,h}$, and c_d are as defined in Definitions 2.3.1 and 2.3.2.

The following remarks and Theorem G.1 characterize the solution to Problem G.2. They are presented here without proof, as they can be directly inferred from the arguments in Wesołowski and Wieczorkowski [45, Proof of Th. 2.1, p. 2215 and Proof of Prop. 2.2, p. 2216].

Remark G.1. Problem G.2 admits an optimal solution for any $p \in \mathcal{P}$.

Remark G.2. Let $(T^*,\, \boldsymbol{x}^*)$ be the solution to Problem G.2 for $p = (\mathcal{H},\, \boldsymbol{N},\, \boldsymbol{S},\, \boldsymbol{\rho},\, n) \in \mathcal{P}$, and define

$$n_{max} := \sum_{d\in\delta(\mathcal{H})} \frac{\big(\sum_{h\in\eta_d(\mathcal{H})} N_{d,h} S_{d,h}\big)^2}{\sum_{h\in\eta_d(\mathcal{H})} N_{d,h} S_{d,h}^2}. \tag{G.6}$$

Then, the following implications hold:

$$\begin{aligned} n \in (0,\, n_{max}) \quad &\Longrightarrow \quad T^* > 0, \\ n \in \Big[n_{max},\, \sum_{(d,h)\in\mathcal{H}} N_{d,h}\Big] \quad &\Longrightarrow \quad T^* \le 0. \end{aligned} \tag{G.7}$$

Remark G.2 implies that in Problem G.1, if $n \in (0, n_{max})$, the constraint (G.2c) can be omitted without changing the optimal solution. Indeed, this insight was utilized by Wesołowski and Wieczorkowski [45], where the optimal solution was derived via the method of Lagrange multipliers considering only equality constraints.

We recall that in the CPDA problem, due to the upper-bound constraints (3.2c), the optimal solution $(T^*, \boldsymbol{x}^*)$ satisfies $T^* \geq 0$ for any $n \in \left(0, \sum_{(d,h)\in\mathcal{H}} N_{d,h}\right]$; see Corollary 3.1.1.

Theorem G.1. *Let the operator* Eigen *and the function s be defined as in Section 3.2.1. Given model $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$, a point $(T^*, \boldsymbol{x}^*) \in \mathbb{R}^{1+|\mathcal{H}|}$ is the optimal solution to Problem G.2 if*

$$\begin{aligned} T^* &= \lambda^*, \\ x^*_{d,h} &= s(\emptyset, \boldsymbol{v}^*, d \mid p)\, A_{d,h}(p), \qquad (d,h) \in \mathcal{H}, \end{aligned} \tag{G.8}$$

where $(\lambda^*, \boldsymbol{v}^*) := \mathrm{Eigen}(\emptyset \mid p)$.

Remark G.3. The `DCA`$(p, \emptyset)$ algorithm computes the optimal solution to Problem G.2 for any $p \in \mathcal{P}$.

Proof of Remark G.3. This statement follows directly from Theorem G.1 and the definition of the `DCA` algorithm. □

Remark G.3 justifies the use of a broader requirement for n in the definition of `DCA`. Specifically, as a base-case algorithm within the `RDCA` recursive framework, `DCA` must be capable of solving Problem G.2 for any $p = (\mathcal{H}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{\rho}, n) \in \mathcal{P}$ (recall (2.4e)).

Bibliography

[1] Avriel, M. [2003]. *Nonlinear Programming: Analysis and Methods*, Dover Publications, Inc. ISBN10: 0-486-43227-0 (pbk.), ISBN13: 978-0-486-43227-4 (pbk.).

[2] Baillargeon, S. and Rivest, L.-P. [2011]. The construction of stratified designs in R with the package `stratification`, *Survey Methodology*, 37(1), pp. 53–65.
`https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201100111447`

[3] Barcaroli, G. [2014]. `SamplingStrata`: An R Package for the Optimization of Stratified Sampling, *Journal of Statistical Software*, 61(4), pp. 1–24.
`https://www.jstatsoft.org/index.php/jss/article/view/v061i04`

[4] Bertsekas, D. P. [2016]. *Nonlinear Programming*, 3rd edn, Athena Scientific. ISBN10: 1-886529-05-1, ISBN13: 978-1-886529-05-2.

[5] Boyd, S. and Vandenberghe, L. [2004]. *Convex Optimization*, Cambridge University Press, Cambridge. ISBN: 0-521-83378-7, ISBN: 978-0-521-83378-3 (hardback).
`http://www.cambridge.org/9780521833783`

[6] Brito, J., Silva, P. and Veiga, T. [2017]. `stratbr`*: Optimal Stratification in Stratified Sampling.* R package version 1.2.
`https://CRAN.R-project.org/package=stratbr`

[7] Burgard, J. P. and Münnich, R. T. [2012]. Modelling over and undercounts for design-based Monte Carlo studies in small area estimation: An application to the German register-assisted census, *Computational Statistics & Data Analysis*, 56, pp. 2856–2863.
`https://www.sciencedirect.com/science/article/pii/S0167947310004305`

[8] Choudhry, G. H., Hidiroglou, M. and Rao, J. [2012]. On sample allocation for efficient domain estimation, *Survey Methodology*, 38(1), pp. 23–29.

https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201200111682

[9] Cochran, W. G. [1977]. *Sampling Techniques*, 3rd edn, John Wiley & Sons, New York.

[10] Cont, R. and Heidari, M. [2014]. Optimal rounding under integer constraints, , . https://arxiv.org/abs/1501.00014

[11] Dalenius, T. [1953]. The multivariate sampling problem, *Skandinavisk Aktuarietidskrift*, 36, pp. 92–102.

[12] de Moura Brito, J. A., do Nascimento Silva, P. L., Silva Semaan, G. and Maculan, N. [2015]. Integer programming formulations applied to optimal allocation in stratified sampling, *Survey Methodology*, 41(2), pp. 427–442. https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201500214249

[13] European Commission and Eurostat [2013]. *Handbook on precision requirements and variance estimation for ESS households surveys – 2013 edition*, Publications Office of the European Union. https://data.europa.eu/doi/10.2785/13579

[14] Friedrich, U., Münnich, R., de Vries, S. and Wagner, M. [2015]. Fast integer-valued algorithms for optimal allocations under constraints in stratified sampling, *Computational Statistics & Data Analysis*, 92, pp. 1–12. https://www.sciencedirect.com/science/article/pii/S0167947315001413

[15] Gentle, J. E. [2017]. *Matrix Algebra*, 2nd edn, Springer International Publishing AG, Cham, Switzerland. ISBN: 978-3-319-64866-8, ISBN: 978-3-319-64867-5 (eBook).

[16] Gunning, P. and Horgan, J. M. [2004]. A New Algorithm for the Construction of Stratum Boundaries in Skewed Populations, *Survey Methodology*, 30(2), pp. 159–166. https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X20040027749

[17] Hartley, H. O. [1965]. Multiple purpose optimal allocation in stratified sampling, *Proceedings of the Social Statistics Section*, 8, pp. 258–261. http://www.asasrms.org/Proceedings/y1965/Multiple%20Purpose%20Optimum%20Allocation%20In%20Stratified%20Sampling.pdf

[18] Hiriart-Urruty, J.-B. and Lemaréchal, C. [1993]. *Convex Analysis and Minimization Algorithms I*, second corrected printing 1996 edn, Springer-Verlag, Berlin. ISBN: 978-3-642-08161-3, ISBN: 978-3-662-02796-7 (eBook).

[19] Hoare, C. A. R. [1969]. An axiomatic basis for computer programming, *Commun.*

ACM, 12(10), pp. 576–580.
https://doi.org/10.1145/363235.363259

[20] Kato, T. [1982]. *A Short Introduction to Perturbation Theory for Linear Operators*, softcover reprint of the hardcover 1 st edn, Springer-Verlag New York Inc., New York, NY. ISBN-13: 978-1-4612-5702-8, e-ISBN-13: 978-1-4612-5700-4.

[21] Khan, M. G. M., Nand, N. and Ahmad, N. [2008]. Determining the optimum strata boundary points using dynamic programming, *Survey Methodology*, 34(2), pp. 205–214.
https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X200800210761

[22] Khan, M. G. M. and Wesołowski, J. [2019]. Neyman-type sample allocation for domains-efficient estimation in multistage sampling, *AStA Advances in Statistical Analysis*, 103(4), pp. 563–592.
https://doi.org/10.1007/s10182-018-00340-2

[23] Kish, L. [1965]. *Survey Sampling*, John Wiley & Sons, Inc., New York. ISBN: 0-471-48900-X.

[24] Kozak, M., Zieliński, A. and Singh, S. [2008]. Stratified two-stage sampling in domains: Sample allocation between domains, strata, and sampling stages, *Statistics & Probability Letters*, 78(8), pp. 970–974.
https://www.sciencedirect.com/science/article/pii/S0167715207003495

[25] Lednicki, B. and Wieczorkowski, R. [2003]. Optimal Stratification and Sample Allocation between Subpopulations and Strata, *Statistics in Transition*, 6(2), pp. 287–305.
https://stat.gov.pl/download/gfx/portalinformacyjny/en/defaultstronaopisowa/3432/1/1/sit_volume_4-7.zip

[26] Lohr, S. L. [2021]. *Sampling*, 3rd edn, Chapman and Hall/CRC., New York. e-ISBN: 978-0-429-29889-9.
https://doi.org/10.1201/9780429298899

[27] Malla, K. R. and Bajracharya, P. M. [2019]. A Study of the Relationship between Convex Sets and Connected Sets, *Journal of Advanced Academic Research (JAAR)*, 6(1), pp. 18–28.
https://doi.org/10.3126/jaar.v6i1.35311

[28] Meyer, C. D. [2000]. *Matrix Analysis and Applied Linear Algebra*, Society for

Industrial and Applied Mathematics. ISBN:0-89871-454-0.

[29] Mordukhovich, B. S. and Nam, N. M. [2014]. *An Easy Path to Convex Analysis and Applications*, Morgan & Claypool. ISBN: 9781627052375 (paperback), ISBN: 9781627052382 (ebook).

[30] Munkres, J. R. [2013]. *Topology (Pearson New International Edition)*, Pearson Education Limited. ISBN10: 1-292-02362-7, ISBN13: 978-1-292-02362-5.

[31] Münnich, R. T., Sachs, E. W. and Wagner, M. [2012]. Numerical solution of optimal allocation problems in stratified sampling under box constraints, *AStA Advances in Statistical Analysis*, 96(3), pp. 435–450.
`https://doi.org/10.1007/s10182-011-0176-z`

[32] Niemiro, W. and Wesołowski, J. [2001]. Fixed Precision Optimal Allocation in Two-Stage Sampling, *Applicationes Mathematicae*, 28(1), pp. 73–82.

[33] Nocedal, J. and Wright, S. J. [2006]. *Numerical Optimization*, Springer New York, NY. Hardcover ISBN: 978-0-387-30303-1, Softcover ISBN: 978-1-4939-3711-0, e-ISBN: 978-0-387-40065-5.
`https://doi.org/10.1007/978-0-387-40065-5`

[34] Pons, M. [2014]. *Real Analysis for the Undergraduate*, Springer New York, New York. ISBN: 978-1-4614-9637-3, ISBN: 978-1-4614-9638-0 (eBook).

[35] R Core Team [2026]. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
`https://www.R-project.org/`

[36] Rockafellar, R. T. [1997]. *Convex Analysis*, Vol. 28 in Princeton Mathematical Series, Princeton University Press, Princeton, New Jersey. ISBN: 0-691-01586-4 (paperback), ISBN13: 978-0-691-1586-6 (paperback).

[37] Rudin, W. [1976]. *Principles of Mathematical Analysis*, 3rd edn, McGraw Hill. ISBN: 0-07-054235-X.

[38] Schaich, E. and Münnich, R. [2001]. Zum Allokationsproblem bei mehreren Untersuchungsvariablen, *Allgemeines Statistisches Archiv*, 77, pp. 390–405.

[39] Singh, T. B. [2019]. *Introduction to Topology*, Springer Nature Singapore Pte Ltd., Sinagpore. ISBN: 978-981-13-6953-7, ISBN: 978-981-13-6954-4 (eBook).

[40] Sohrab, H. H. [2014]. *Basic Real Analysis*, 2nd edn, Birkhäuser New York, New York. ISBN: 978-1-4939-1840-9, ISBN: 978-1-4939-1841-6 (eBook).

[41] Stenger, H. and Gabler, S. [2005]. Combining random sampling and census strategies - Justification of inclusion probabilities equal to 1, *Metrika*, 61(2), pp. 137–156.
https://doi.org/10.1007/s001840400328

[42] Särndal, C.-E., Swensson, B. and Wretman, J. [1992]. *Model Assisted Survey Sampling*, Springer New York, NY.

[43] Valliant, R., Dever, J. A. and Kreuter, F. [2018]. *Practical Tools for Designing and Weighting Survey Samples*, 2nd edn, Springer Cham.

[44] Wesołowski, J. [2019]. Multi-domain Neyman-Tchuprov optimal allocation, *Statistics in Transition new series*, 20(4), pp. 1–12.
https://doi.org/10.21307/stattrans-2019-031

[45] Wesołowski, J. and Wieczorkowski, R. [2017]. An eigenproblem approach to optimal equal-precision sample allocation in subpopulations, *Communications in Statistics - Theory and Methods*, 46(5), pp. 2212–2231.
https://doi.org/10.1080/03610926.2015.1040501

[46] Wesołowski, J., Wieczorkowski, R. and Wójciak, W. [2022]. Optimality of the Recursive Neyman Allocation, *Journal of Survey Statistics and Methodology*, 10(5), pp. 1263–1275.
https://academic.oup.com/jssam/article-pdf/10/5/1263/46878255/smab018.pdf

[47] Wesołowski, J., Wieczorkowski, R. and Wójciak, W. [2024]. Recursive Neyman algorithm for optimum sample allocation under box constraints on sample sizes in strata, *Survey Methodology*, 50(2), pp. 487–511.
http://www.statcan.gc.ca/pub/12-001-x/2024002/article/00003-eng.pdf

[48] Willems, F. [2025]. *A Framework for Multiobjective and Uncertain Resource Allocation Problems in Survey Sampling based on Conic Optimization*, PhD thesis, Universität Trier.
https://ubt.opus.hbz-nrw.de/frontdoor/index/index/docId/2592

[49] Wright, T. [2017]. Exact optimal sample allocation: More efficient than Neyman, *Statistics & Probability Letters*, 129, pp. 50–57.
https://www.sciencedirect.com/science/article/pii/S0167715217301657

[50] Wright, T. [2020]. A general exact optimal sample allocation algorithm: With bounded cost and bounded sample sizes, *Statistics & Probability Letters*, 165,

pp. 108829.
https://www.sciencedirect.com/science/article/pii/S0167715220301322

[51] Wójciak, W. [2026]. stratallo*: Optimum Sample Allocation in Stratified Sampling.* R package version 3.0.0.
https://CRAN.R-project.org/package=stratallo

[52] Wójciak, W., Wesołowski, J. and Wieczorkowski, R. [2026]. R package stratallo - source code.
https://github.com/wwojciech/stratallo

[53] Yates, F. [1971]. *Sampling Methods for Censuses and Surveys*, 3rd edn, Griffin and Company, Ltd., London.